

Petroleum-Resource Appraisal and Discovery Rate Forecasting in Partially Explored Regions

- An Application to the Denver basin
- Mathematical Foundations
- An Application to Supply Modeling

GEOLOGICAL SURVEY PROFESSIONAL PAPER 1138-A, B, C



Petroleum-Resource Appraisal and Discovery Rate Forecasting in Partially Explored Regions

An Application to the Denver basin

By L. J. DREW, J. H. SCHUENEMEYER, *and* D. H. ROOT

Mathematical Foundations

By D. H. ROOT *and* J. H. SCHUENEMEYER

An Application to Supply Modeling

By E. D. ATTANASI, L. J. DREW, *and* J. H. SCHUENEMEYER

GEOLOGICAL SURVEY PROFESSIONAL PAPER 1138-A, B, C



UNITED STATES DEPARTMENT OF THE INTERIOR

CECIL D. ANDRUS, *Secretary*

GEOLOGICAL SURVEY

H. William Menard, *Director*

Library of Congress Cataloging in Publication Data

Petroleum resource appraisal and discovery rate forecasting in partially explored regions

Geological Survey Professional Paper 1138

CONTENTS: A. Drew, L. J., Schuenemeyer, J. H., and Root, D. H., An application to the Denver basin

B. Root, D. H., and Schuenemeyer, J. H., Mathematical foundations

C. Attanasi, B. E. D., Drew, L. J., and Schuenemeyer, J. H., An application to supply modeling

Supt. of Docs. No.: I 19.16:1138

1. Prospecting—Denver Basin, Colo. and Wyo.—Mathematical models. 2. Petroleum—Denver Basin, Colo. and Wyo.—Mathematical models. I. Drew, Lawrence J. II. Series: United States. Geological Survey. Professional Paper 1138

TN271.P4P47

622'.18282'09788

80-607018

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Petroleum-Resource Appraisal and Discovery Rate Forecasting in Partially Explored Regions—

An Application to the Denver basin

By L. J. DREW, J. H. SCHUENEMEYER, and D. H. ROOT

GEOLOGICAL SURVEY PROFESSIONAL PAPER 1138-A

*A model of the discovery process that can be used to predict
the size distribution of future petroleum discoveries
in partially explored basins*



CONTENTS

Abstract	Page A1
Introduction	1
Area of influence of a drill hole	1
Estimation of effective basin area	2
Concept of discovery efficiency	5
Estimation of discovery efficiencies in the Denver basin	5
Predicting future discoveries within a size class	7
Predicting future discoveries across size classes	8
Discussion	10
References cited	11

ILLUSTRATIONS

FIGURES 1-11. Graphs showing:	Page
1. Average area exhausted per well (including both development and exploratory) for different target sizes and number of wells	A2
2. Total area exhausted versus target size for selected numbers of wells	3
3. Area exhausted and effective basin size versus cumulative wells for 4.3-mi ² targets	4
4. Cumulative fraction of targets discovered versus fraction of basin exhausted for two different hypothesized efficiencies	5
5. Possible anomalous discovery rate that initially appears to have a discovery efficiency less than 1	6
6. Computed discovery efficiencies for classes 9 through 15 for years 1963-69	6
7. Cumulative discoveries versus area exhausted for class size 9 (256,000 to 512,000 bbl)	6
8. Cumulative discoveries versus area exhausted for class size 10 (512,000 to 1,024,000 bbl)	7
9. Cumulative discoveries versus area exhausted for class size 12 (2,048,000 to 4,096,000 bbl)	8
10. Cumulative discoveries versus area exhausted for class size 13 (4,096,000 to 8,192,000 bbl)	9
11. Seventeen successive predictions of the total volume of petroleum to have been discovered by the end of 1969 in classes 9 through 15	9
12. Histogram comparing actual versus predicted petroleum discovered within size classes 9 through 15 for discoveries made between January 1, 1955, and December 31, 1969, based on data to December 31, 1954	10
13. Graph showing 22 successive predictions of the total volume of petroleum to have been discovered by the end of 1974 in classes 9 through 15	10
14. Histogram comparing actual versus predicted petroleum discovered within size classes 9 through 15 for discoveries made between January 1, 1956, and December 31, 1974, based on data to December 31, 1955	10

TABLES

TABLE	Page
1. Conversion factors used in computation for development well conversions	A3
2. Expected number of discoveries per 1,000 square miles searched for efficiencies of 1 and 2.5	5
3. Sample calculations for future discoveries in size class 12	8

PETROLEUM-RESOURCE APPRAISAL AND DISCOVERY RATE FORECASTING IN PARTIALLY EXPLORED REGIONS— AN APPLICATION TO THE DENVER BASIN

By L. J. DREW, J. H. SCHUENEMEYER, and D. H. ROOT

ABSTRACT

A model of the discovery process can be used to predict the size distribution of future petroleum discoveries in partially explored basins. The parameters of the model are estimated directly from the historical drilling record, rather than being determined by assumptions or analogies. The model is based on the concept of the area of influence of a drill hole, which states that the area of a basin exhausted by a drill hole varies with the size and shape of targets in the basin and with the density of previously drilled wells. It also uses the concept of discovery efficiency, which measures the rate of discovery within several classes of deposit size. The model was tested using 25 years of historical exploration data (1949–74) from the Denver basin. From the trend in the discovery rate (the number of discoveries per unit area exhausted), the discovery efficiencies in each class of deposit size were estimated. Using pre-1956 discovery and drilling data, the model accurately predicted the size distribution of discoveries for the 1956–74 period.

INTRODUCTION

Before the cost of petroleum obtained from future discoveries can be determined, a model of the discovery process must be used to predict not only the total volume of petroleum to be discovered with any given level of exploratory drilling but also the size distribution of these discoveries. This detailed information is required because, as the discovery process proceeds, the incremental size distribution of discoveries progressively contains a higher percentage of smaller deposits. Cost rises not only because progressively smaller volumes of petroleum are found per unit of exploratory effort but also because smaller deposits usually have higher per unit development costs.

The purpose of this investigation was to develop a model of the discovery process that will predict the incremental size distribution of future discoveries in any partially explored region as a function of exploratory drilling effort. In the construction of this model, a structure was specified whose parameters can be estimated directly from historical drilling and discovery data rather than by subjective judgment or the use of an analogy. The number of discoveries predicted in each size class is determined by the value of two pa-

rameters. The first parameter, which is commonly estimated for all size classes, is the effective basin area, which is that region within which drillers are willing to site wells. The second parameter, which is uniquely estimated for each size class, is the discovery efficiency parameter. This parameter measures the effectiveness of exploratory drilling relative to the outcome of a random search process. The higher the value of the discovery efficiency parameter in any given size class, the earlier most of the deposits in that class will be discovered.

The model of the discovery process described in this report is related to the model proposed by Arps and Roberts (1958) but differs in several significant aspects. First, the discovery efficiencies are estimated within each deposit size class rather than assumed to be a single efficiency for all size classes as determined by subjective judgment. Second, the effective basin area is estimated from discovery time series data, in contrast to the approach taken by Arps and Roberts in which the boundaries of the region to be explored were selected by expert judgment.

In the model used here, it was possible to estimate these parameters directly from discovery time series data by using the concept of the area of influence of a drill hole (Singer and Drew, 1976). Using this approach, it was possible to predict accurately the number of deposits to be discovered within a group of size classes in the Denver basin from only a small amount of historical data. The Denver basin was chosen to test this model because it contains only one major producing interval (the D and J members of the Cretaceous Dakota Sandstone), and so a reliable data base could be constructed within a reasonable period of time (Schuenemeyer and Drew, 1977b).

AREA OF INFLUENCE OF A DRILL HOLE

The area exhausted by a drill hole varies with the size of the targets that can occur in the basin. A basin may be fully explored with respect to large targets but

only 10 percent explored with respect to small targets. The shape of the targets also has an effect; for example, a pattern of exploratory and development holes may have fully searched a basin for circular targets having areas of 2 mi², although irregularly shaped targets also having an area of 2 mi² are still undiscovered.

The amount of area exhausted by any given hole can range from a maximum of the area of the target being considered when the areas of influence of any previously drilled holes do not overlap to a minimum of zero when all points within the area of influence of the hole have been exhausted by previously drilled holes. Singer and Drew (1976) and Root and Schuenemeyer (1980) discuss this concept in detail, and Singer (1976) and Schuenemeyer and Drew (1977a) describe the computations. The degree to which the Cretaceous Dakota Sandstone interval in the Denver basin was physically exhausted by the end of each year within the period 1949–74 was calculated for targets up to 4.3 mi² in area (fig. 1). The levels of exhaustion shown are for elliptical targets with an axial ratio of 0.38, the mean axial ratio of deposits discovered during 1949–74. For example, the 1,096 wells drilled through the end of 1952 exhausted on the average 0.545 mi² per well with respect to targets 1.09 mi² in area. By the end of 1954, this average declined 12.4 percent to 0.478 mi² per well. This decline in the effective exhaustiveness of wells is a consequence of the progressive crowding of wells. For larger targets this crowding is even more severe. For example, for targets 4.3 mi² in area, the corresponding area exhausted per well declines by 26 percent from 1.71 to 1.26 mi² per well during 1952–54.

In 1969 the Union Pacific Railroad released for exploration approximately 4,400 mi² of virtually unexplored land in the basin (Oil and Gas Journal, 1969). The release and exploration of this large block of land during subsequent years are reflected in figure 1 by a slight increase in the level of the 1974 curve over the 1968 curve. This increase is a result of the wider spacing of holes, which normally occurs during the early phase of exploration of any region.

ESTIMATION OF EFFECTIVE BASIN AREA

The effective basin area is defined as that part of a total basin within which drillers are willing to site wells. Given this definition, there is no reason to assume that the area searched will have any particular shape. One basin may be more or less evenly explored from rim to rim. In others, many wells may be clustered in small parts of the total basin area. In such basins, the effective basin area may not even be a con-

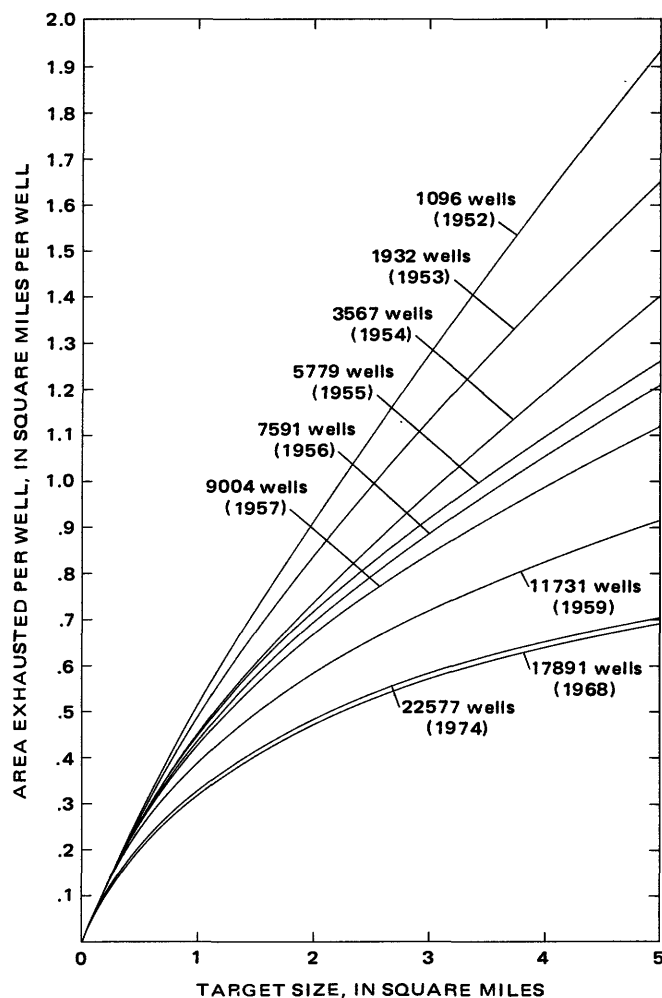


FIGURE 1.—Average area exhausted per well (including both development and exploratory) for different target sizes and number of wells.

nected region.¹ Using the concept of area of influence of a drill hole to measure the crowding of exploratory and development drilling, a method has been developed to estimate the effective basin area.

As a simplified example of this method, suppose that an irregular area is marked off on the floor, and then disks are placed at random inside the area so that they have as much chance of landing one place as another and they may overlap. After several disks have been placed, the boundaries of the area are erased, but the disks are left on the floor. The problem then is to calculate the area, *B*, inside the boundaries that have now been erased. Suppose the disks each have area "*a*." In the current analysis, the centers of the disks are the locations of the wells. The disk area is taken to be the

¹A region is connected if and only if any two points in the region can be joined by a curve line that does not go outside of the region (region is in one piece).

area of the largest target under consideration, and B is the effective basin area. The average target area corresponding to class 15, the largest deposit size under consideration, is 4.3 mi². Let the function $A(n)$ measure the area covered by the first n disks. The value of this function will then be $A(n)=na$ if none of the disks overlaps and $A(n) < na$ when they do. Assume, furthermore, that the disks are small compared to B , the effective basin area. It can then be shown that even though $A(n)$ is a random variable, it grows in a regular fashion as n increases (Root and Schuenemeyer, 1980). On the average $A(n)$ grows according to a negative exponential:

$$A(n) = B(1 - e^{-na/B}). \quad (1)$$

If an area $A(n)$ has already been exhausted, then the additional area that will be exhausted by m additional wells sited at random within B is given by

$$A(n+m) - A(n) = (B - A(n))(1 - e^{-ma/B}). \quad (2)$$

A proof of these identities is provided by Root and Schuenemeyer (1980).

In general, any arbitrary shape having the same area can be used without changing the rate of growth of $A(n)$. In the following calculations, the targets are assumed to be elliptical. Given the area exhausted by the exploratory and development wells drilled through time (fig. 2), an estimate of the effective basin area for the largest class can be obtained in the case for which all wells are sited at random by solving equation 1 for B . In practice, however, development wells are not drilled at random but instead in a closely spaced regular pattern. In the Denver basin, most development wells have been drilled one-quarter mile apart. As a consequence, a development well cannot have the same average capacity to exhaust a region as an exploratory well. We therefore treat a development well as a fraction of an exploratory well. At any time, then, the equivalent number of exploratory wells that have been drilled is specified by

$$\hat{n} = W + Dd(a) \quad (3)$$

where

- \hat{n} = equivalent number of exploratory wells,
 - W = number of true exploratory wells,
 - D = number of development wells, and
 - $d(a)$ = conversion factor for development wells.
- a = target size

The conversion factor, $d(a)$, is a function of the target size, the target shape, and the spacing of development wells. The value of this factor for very small targets is

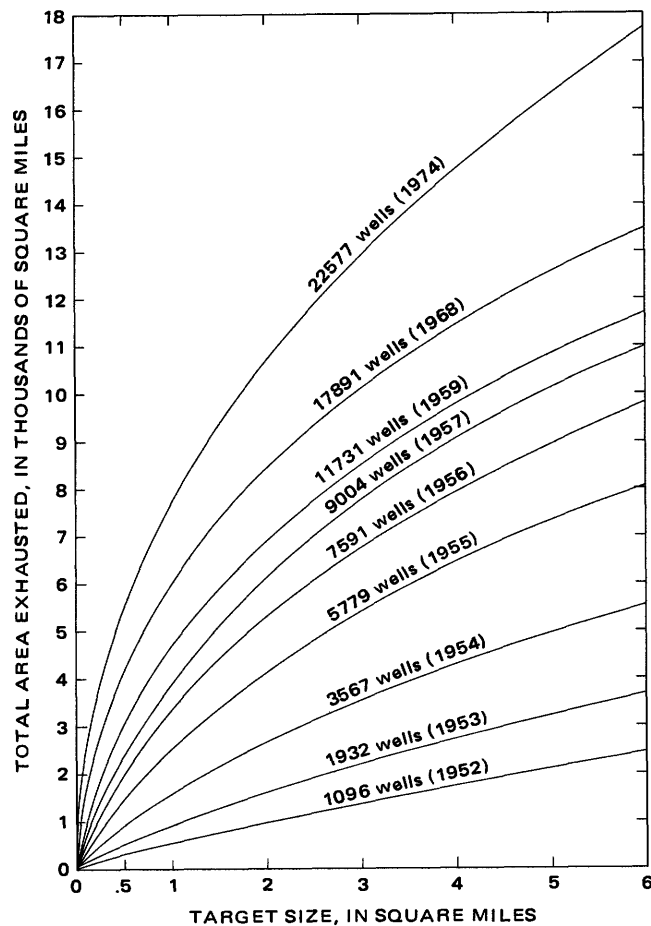


FIGURE 2.—Total area exhausted versus target size for selected numbers of wells.

equal to one, and, as the target size increases, $d(a)$ decreases toward zero. If the spacing between holes is one-quarter mile and the axial ratio is 0.38, then targets with an area not greater than 0.0187 mi² will have $d(a)=1$. This factor is chosen so that two target-like ellipses having the same orientation and centers one-quarter mile apart will have an area of overlap, averaged over all orientations of $(1-d(a))a$. Conversion factors for various sizes of ellipses with axial ratio 0.38 are presented in table 1.

TABLE 1.—Conversion factors used in computation for development well conversions

Class	Size (10 ³ barrels)	Conversion factor $d(a)$	Average producing area (mi ²)
9	256– 512	0.54	0.339
10	512– 1,024	.45	.513
11	1,024– 2,048	.37	.753
12	2,048– 4,096	.31	1.091
13	4,096– 8,192	.25	1.670
14	8,192– 16,384	.23	2.070
15	16,384– 32,768	.16	4.300

Given the number and location of exploratory and development holes drilled, we can now solve for the effective basin area using equation 4,

$$A(\hat{n}) = B(1 - e^{-\hat{n}a/B}). \quad (4)$$

For estimating the effective basin area, \hat{n} is computed using $d(a=4.3)=0.16$, the conversion factor for the largest target size.

If an area $A(\hat{n})$ has already been exhausted within an effective basin area B , then the additional area to be exhausted with respect to a target of area a by an additional number of equivalent exploratory wells \hat{m} is

$$A(\hat{n} + \hat{m}) - A(\hat{n}) = (B - A(\hat{n}))(1 - e^{-\hat{m}a/B}). \quad (5)$$

The estimate \hat{m} is obtained in a manner similar to that of \hat{n} , except that the future ratio of the number of development holes to exploration holes must be assumed. We assumed that the future ratio would be the same as the past ratio.

With the results of the calculations from equation 4,

both the actual area exhausted and the estimated effective basin area can be graphed as a function of cumulative wells drilled (fig. 3). The estimated effective basin area increased rapidly through the drilling of the first 9,000 wells (1949–1957). This rapid rate of growth implies that the drillers rapidly expanded their views of where deposits could be discovered in the basin. During the following 10 years, when an additional 8,300 wells were drilled, the effective area of the basin grew very little, which implies that exploration during this period was being confined, for the most part, to the region established by 1957. From 1969 through 1973, the effective basin area again began to grow more rapidly. This growth was caused by the release in 1969 of approximately 4,400 mi² in the Denver basin that had been held by the Union Pacific Railroad. The crowding that resulted in the change in slope in 1957 recurred in 1973—implying that the area to be searched was beginning to be reestablished. It is reasonable to assume that, had such a large region not been withheld from exploration, the effective basin area curves would have shown only one change in slope, from growth to stabilization.

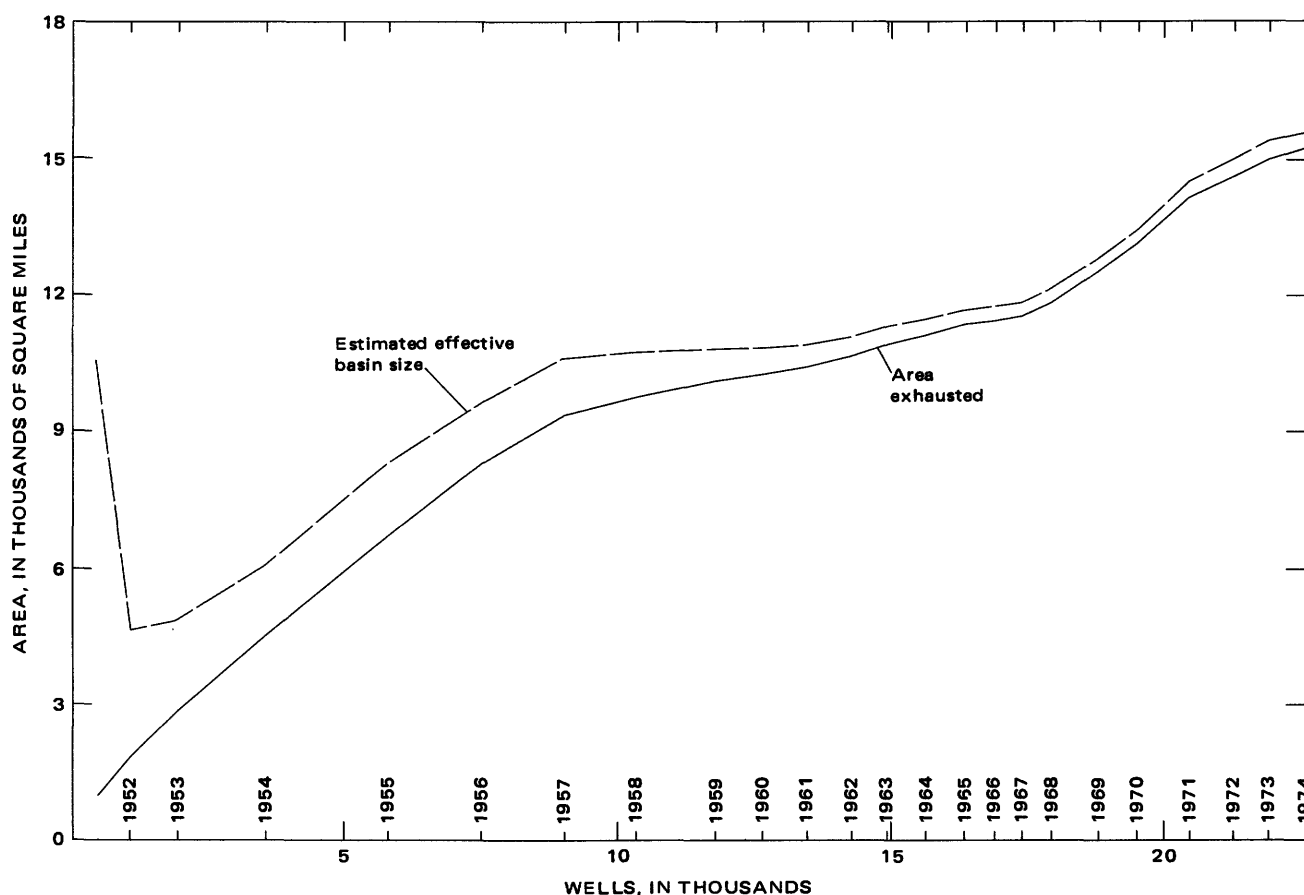


FIGURE 3.—Area exhausted and effective basin size versus cumulative wells for 4.3-mi² targets.

CONCEPT OF DISCOVERY EFFICIENCY

Historical exploration records show that as exploration proceeds, the rate of discovery declines; that is, exploratory drilling begins in the more productive parts of a basin and later moves to the less productive parts of that basin. The concept of discovery efficiency is used to measure this phenomenon. A discovery efficiency of 1 is used as a reference point that defines the outcomes of a random search process. In such a search, the number of deposits discovered in each size class per unit area exhausted remains constant throughout the total exploration history of one region. For example, if the ultimate effective basin area were 10,000 mi² and 100 deposits existed in a certain size class at the start of the search process, we would expect 10 deposits to be discovered with the exhaustion of each 1,000 mi² of the region.

If the exploration process were better than random, we would expect a larger number of deposits to be found during the earlier stages of exploration. For example, if the discovery efficiency were 2.5 in the preceding example, the exhaustion of the first 1,000 mi² would produce 23.2 discoveries versus 10 discoveries for the random search. The number of discoveries expected to be made in this example for each 1,000 mi² exhausted is shown in table 2. These values were computed from the following discovery process model:

$$f = 1 - (1 - A/B)^c \quad (6)$$

where

- f = the fraction of the deposits that have been found,
- A = area that has been exhausted,
- B = effective basin area, and
- c = the discovery efficiency.

A graph of the cumulative number of discoveries as a function of area exhausted for two different efficiencies is schematically diagrammed in figure 4.

TABLE 2.—Expected number of discoveries per 1,000 square miles searched for efficiencies of 1 and 2.5

Area searched (mi ²)	Expected number of discoveries per 1,000 mi ² searched	
	Efficiency = 1	Efficiency = 2.5
1,000	10	23.2
2,000	10	19.6
3,000	10	16.2
4,000	10	13.1
5,000	10	10.2
6,000	10	7.6
7,000	10	5.2
8,000	10	3.1
9,000	10	1.5
10,000	10	.3

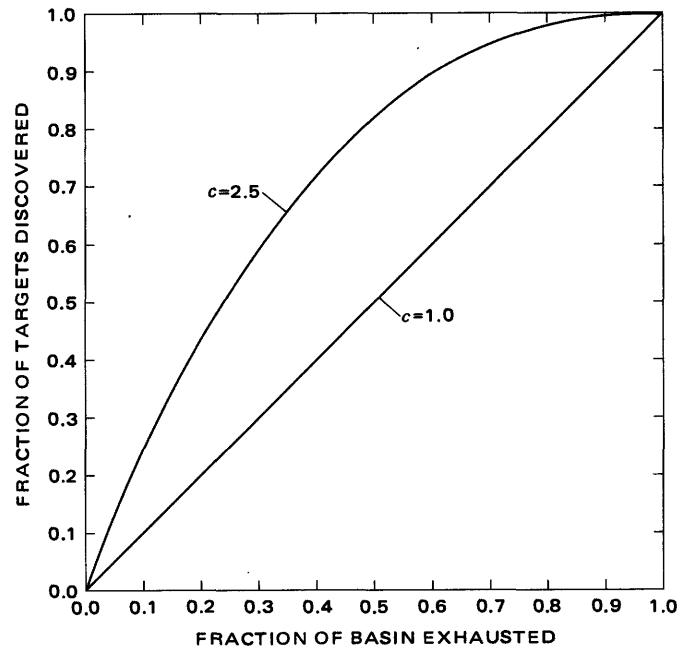


FIGURE 4.—Theoretical graph of cumulative fraction of targets discovered versus fraction of basin exhausted for two different efficiencies.

While it is theoretically possible for a discovery efficiency to be less than 1, this value is of no relevance because it implies a steadily increasing discovery rate through the end of the exploration process; in the long run, drilling would terminate when the discovery rate reaches its peak! In the short run, however, a sudden geologic insight or chance fluctuation could cause the discovery rate to increase for a period of time (fig. 5). If the method for computing efficiencies described below results in an efficiency of less than 1, the efficiency is set equal to 1.

ESTIMATION OF DISCOVERY EFFICIENCIES IN THE DENVER BASIN

From the trend in the discovery rate (the number of discoveries per unit area exhausted), the efficiencies in each class of deposit size were estimated. In the estimation of the efficiencies it is necessary to choose a basin size. The assumption that the entire basin is available for drilling ignores two effects; namely, that exploration tends to follow past discoveries, and drillers tend to explore the shallower parts of a formation before the deeper. These effects combine to produce a greater crowding of wells than would occur with random drilling. This greater crowding results in an underestimation of the area that will ultimately be explored.

We have found a way to adjust the model that compensates for the fact that the area being considered for exploration grows through time. For estimating the

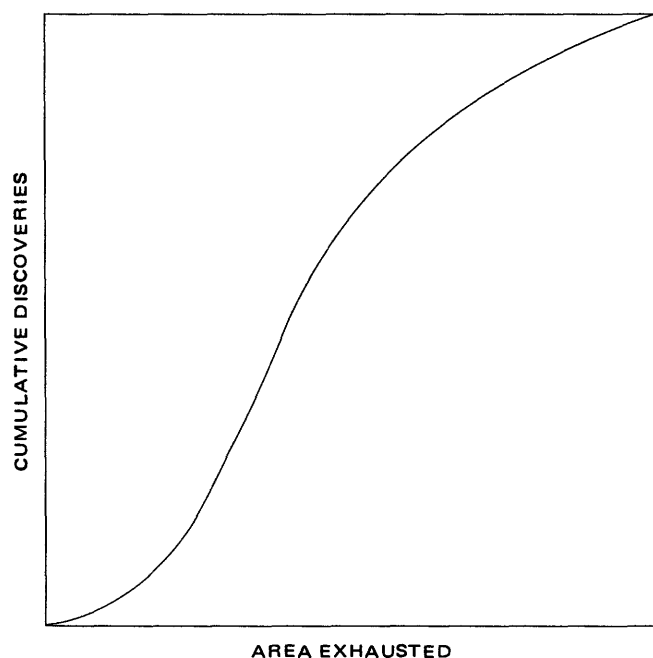


FIGURE 5.—A possible anomalous discovery rate that initially appears to have a discovery efficiency less than 1.

efficiency of exploration for fields of each size class, the basin size used was the basin size estimated for that size class. For estimating future discoveries, the basin size used was that estimated for the largest size class. The theoretical description of the search of an expanding area is still under investigation.

Efficiencies were estimated annually² for the years 1952–69. Seven efficiency estimates (1963–69) for seven deposit size classes are given in figure 6. No trend with time was found within each series of estimates. No estimate was made for class 16 (32.7–65.5 million barrels) because only one deposit was discovered in this, the largest class, and an estimate would be statistically meaningless. Efficiency increases as deposit size increases (fig. 6). Efficiencies less than 1 occurred in classes 9, 10, and 11, but for the purpose of estimating future discoveries, it was assumed that in the long run no efficiency can be less than 1. If the estimated efficiency was less than 1, then for purposes of prediction, it was set equal to 1.

The relation between the discovery efficiency and the shape of the cumulative discovery curve can be seen by comparing the efficiencies estimated in figure 6 with the discovery curves shown in figures 7–10. For example, the discovery efficiency within class 12 was estimated to be approximately 2.6, and the discovery curve has a markedly decreasing slope (fig. 9). An efficiency of this magnitude implies that when half the effective

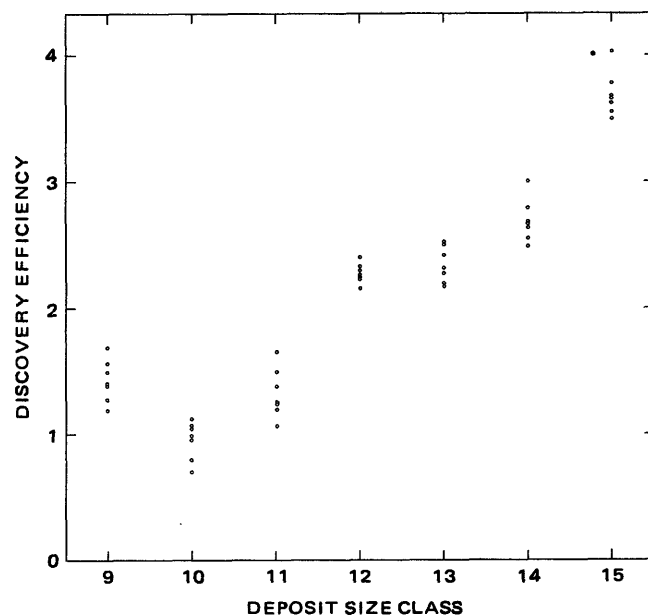


FIGURE 6.—Computed discovery efficiencies for classes 9 through 15 for years 1963–69.

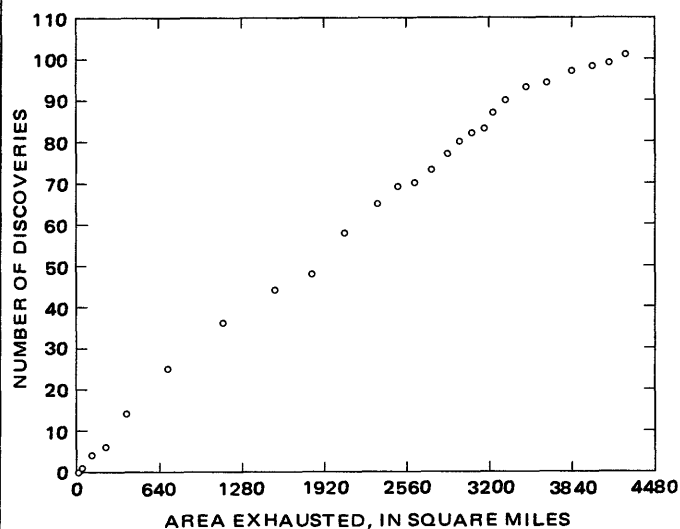


FIGURE 7.—Cumulative discoveries versus area exhausted for class size 9 (256,000 to 512,000 bbl).

basin area has been exhausted, 84 percent of the deposits containing between 2.048 and 4.096 million barrels of recoverable petroleum will have been found. The efficiency of discovery for the deposits in class 9 (0.256 to 0.512 million barrels) was approximately 1.4 (fig. 7). This value means that deposits of this size were found at nearly a uniform rate through 1969, as the nearly straight cumulative discovery curve for this class of deposits shows (fig. 7). The discovery efficiency for the deposits in class 10 (0.512 to 1.024 million barrels) was estimated to be less than 1.0 (fig. 8) be-

²The details of the calculation procedure are presented in Root and Schuenemeyer (1980).

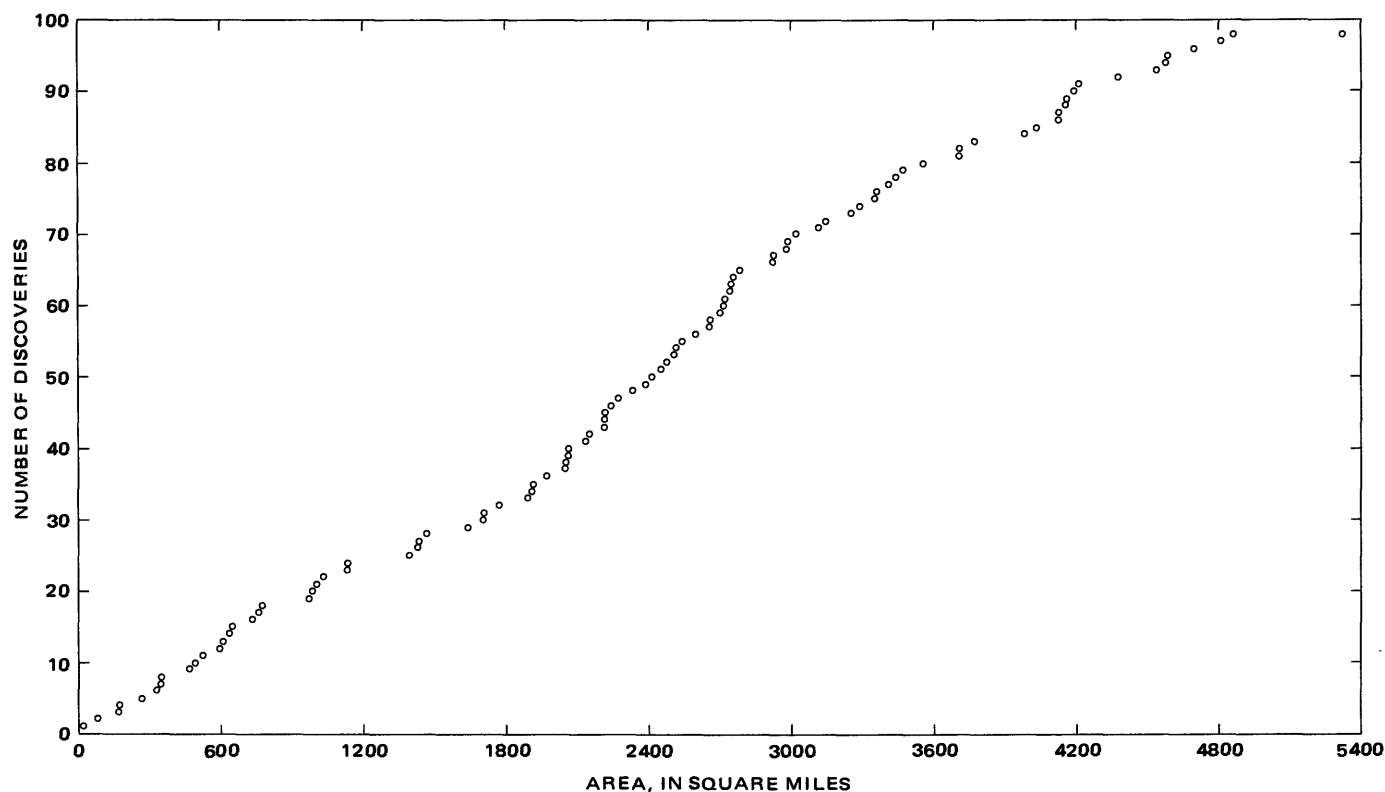


FIGURE 8.—Cumulative discoveries versus area exhausted for class size 10 (512,000 to 1,024,000 bbl).

cause the discovery rate increased rapidly during the exhaustion segment from 1,500 to 3,000 mi². As we have previously discussed, this phenomenon can be only temporary, and, as shown in figure 8, the discovery rate declined during the second half of the exhaustion sequence. No obvious geologic cause could be isolated to explain this temporary increase in the rate of discoveries.

PREDICTING FUTURE DISCOVERIES WITHIN A SIZE CLASS

The model of the discovery process shown in equation 6 was used to predict the ultimate number of deposits to be discovered and the number of deposits to be discovered for a given area exhausted. Each of these predictions is made within a deposit size class. The predictions made for some later segment of exploration are designed to answer the question: How many deposits may one expect to find by drilling a certain number of additional wells, given the previous exploration history?

In order to validate the model given in equation 6, resource appraisal estimates have been made within each deposit size class from the discovery time series for various years beginning in 1952. The last year of historical data used in the prediction is called the pre-

diction year. The number of discoveries estimated from the end of the prediction year until 1969 and until 1974 is compared with the actual number.

Before a forecast can be made, the following estimates are calculated: (1) the ultimate number of deposits within each size class; (2) the area that will be physically exhausted by the number of wells actually drilled between the time the forecast is to be made and the end of 1969 and 1974; and (3) the discovery efficiency.

The number of deposits estimated to have existed originally in each size class is obtained from equation 6. Given the fraction of the effective basin area still remaining to be explored and the corresponding discovery efficiency for each class, the fraction of deposits remaining to be discovered is $(1 - A/B)^{\epsilon}$. The number of deposits then estimated to have occurred originally in each size class is obtained by dividing the number already discovered in a class by $1 - (1 - A/B)^{\epsilon}$. From this calculation we then can construct a resource appraisal for the basin for any segment of the exploration data time series.

The estimate of the basin area for class 15 is obtained from equation 4. Once the basin area has been determined, we then assume that the ratio of exploratory to development wells remains constant. Then we can calculate from equation 5 the additional area to

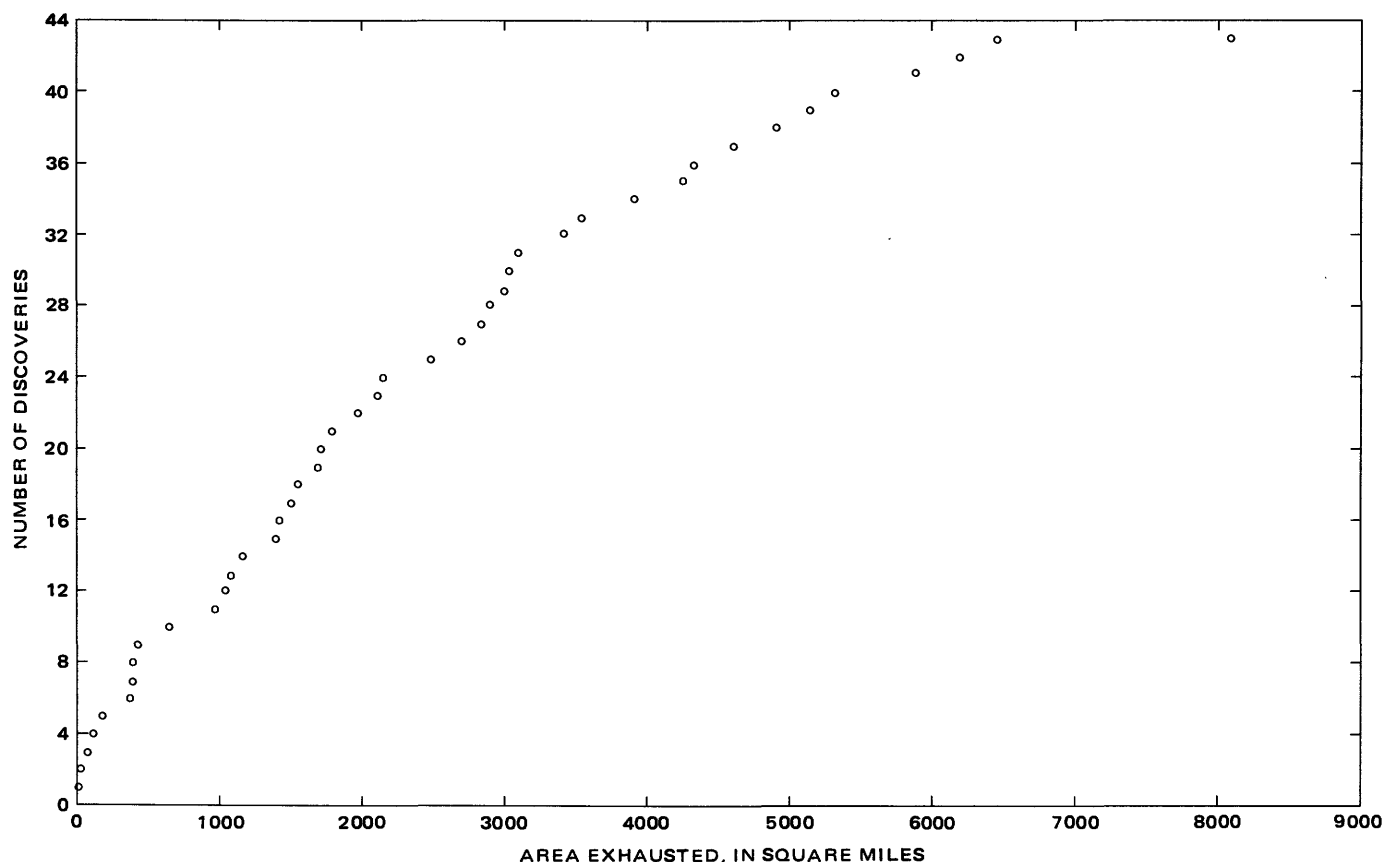


FIGURE 9.—Cumulative discoveries versus area exhausted for class size 12 (2,048,000 to 4,096,000 bbl).

be exhausted for a given number of additional wells. The basin size estimated for deposits of size class 15 is used in this calculation. Using these estimates, the number of discoveries expected within each size class from the end of the prediction year through 1969 and through 1974 are computed.

Sample calculations for future discoveries through 1969 in deposit size class 12 are presented in table 3 for the prediction year 1957. In the computation of adjusted future wells, the number 2 appears as a divisor because the ratio of exploratory to development drilling has been found through time to be approximately equal to 1.

PREDICTION OF FUTURE DISCOVERIES ACROSS SIZE CLASSES

An estimate of the total volume of petroleum to be discovered was obtained by summing the future oil discovered over size classes 9 through 15 (deposits between 0.256 and 32.768 million barrels). This estimate was then compared with the volume actually discovered to test the validity of the prediction. In addition to estimating the total volume of petroleum to be discovered, it is necessary for purposes of economic analysis

TABLE 3.—Sample calculations for future discoveries in size class 12

Basic Data		
Description	Symbol	Value
Area exhausted to December 31, 1957	$A(1957)$	4,165 mi ²
Target area	a	1.09 mi ²
Estimated effective basin size	B	10,624 mi ²
Correction factor for development wells	d	0.31
Discovery efficiency	c	2.577
Number of discoveries to December 31, 1957	N	34
Average volume of recoverable oil		2.863×10^6 bbls
Future wells from January 1, 1958, to December 31, 1969		9,765
Calculations		
Adjusted future wells = $\hat{m} = \frac{9765 + 9765(0.31)}{2} = 6,396$		
$A(1969) = (B - A(1957))(1 - e^{-\hat{m}a/B}) + A(1957) = 7,273$ mi ²		
Unexplored fraction of basin (1957) = $1 - A \frac{(1957)}{B} = 0.608$		
Unexplored fraction of basin (1969) = $1 - A \frac{(1969)}{B} = 0.315$		
Undiscovered fraction of targets (1957) = $(0.608)^c = 0.277$		
Undiscovered fraction of targets (1969) = $(0.315)^c = 0.051$		
Fraction of targets discovered (1958–69) = 0.226		
Estimate of ultimate number of targets = $\frac{N}{1-0.277} = 47.0$		
Estimated number of discoveries (1958–69) = $(47.0) \times (0.226) = 10.6$		
Future oil discovered (1958–69) = $10.6 \times 2.863 \times 10^6 = 30.3 \times 10^6$ bbls		

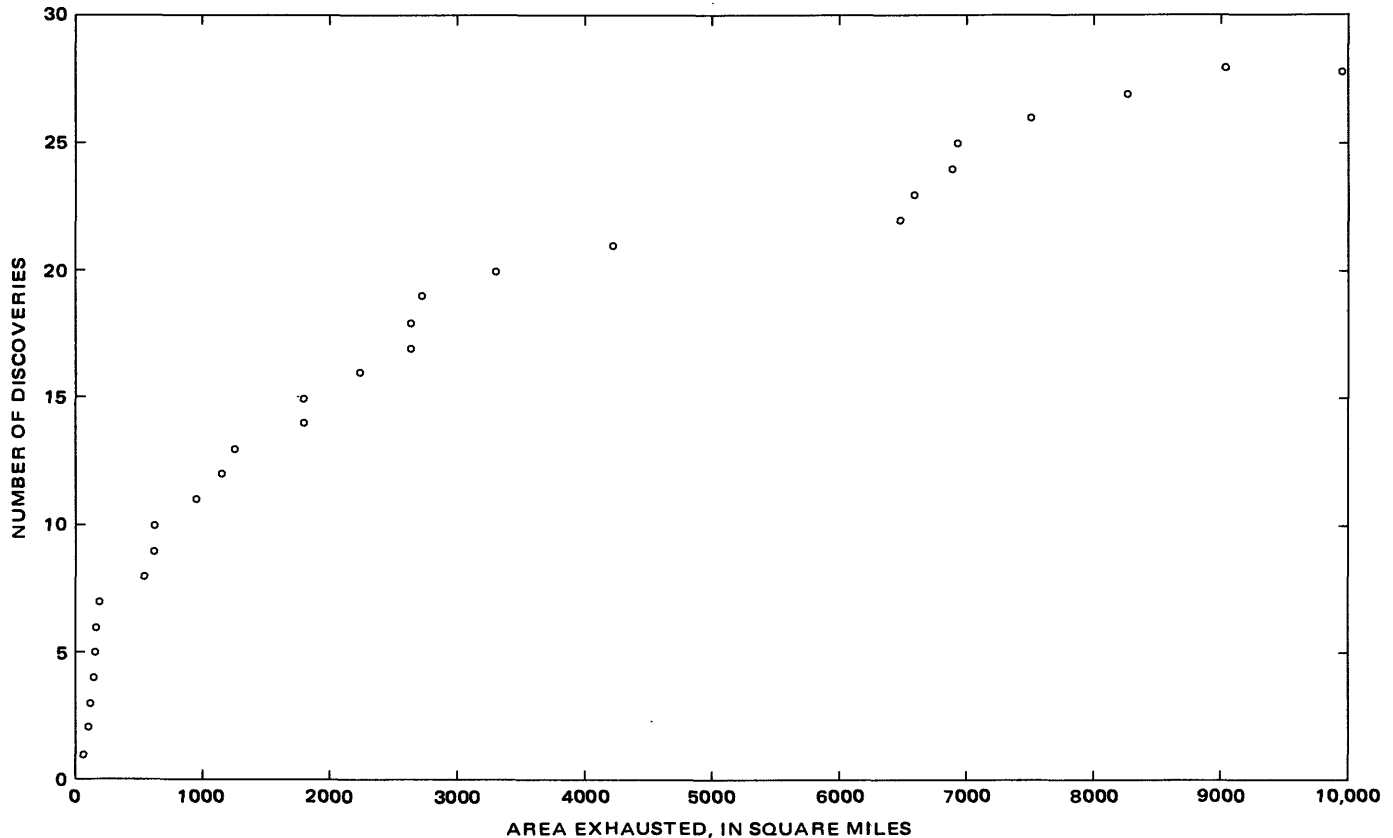


FIGURE 10.—Cumulative discoveries versus area exhausted for class size 13 (4,096,000 to 8,192,000 bbl).

to predict accurately the number of deposits to be discovered in each size class.

The prediction of the total oil to be discovered is the sum of seven quantities, each subject to independent random fluctuations that on the average tend to cancel each other; thus the total is a more stable quantity. On the average, estimated efficiencies tend to increase with deposit size class (fig. 6). Deviations from this trend are attributed to random fluctuations.

The prediction of the total amount of oil to be discovered by drilling 18,762 wells changes as the initial drilling and discovery sequence on which it is based changes (fig. 11). Thus, using only 5 years of data (1949–54), or 19 percent of the wells drilled through 1969, the basin size and efficiencies are sufficiently well determined to permit accurate estimation of the oil to be discovered by the next 15,195 wells. In order to use this estimate for economic calculations, it is necessary to be able to predict the size distribution of deposits. Predicted and actual future discoveries agree closely across size classes using 1954 as the prediction year (fig. 12). This agreement is particularly good in the larger, more important size classes.

The calculations made through 1969 were repeated for the 22,577 wells drilled through 1974 (figs. 13 and 14). During the 5-year period beginning in 1969, the

land released by the Union Pacific Railroad was explored. The late release of this much land might have caused the model slightly to underestimate the amount of oil to have been discovered through the end of 1974.

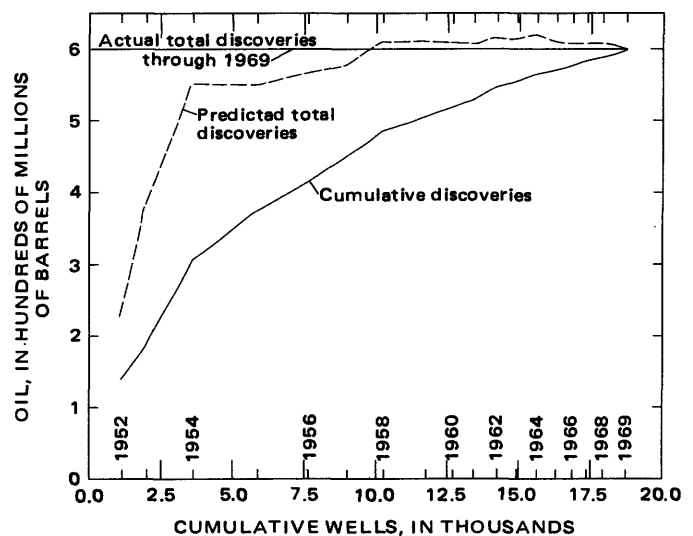


FIGURE 11.—Seventeen successive predictions of the total volume of petroleum to have been discovered by the end of 1969 in classes 9 through 15.

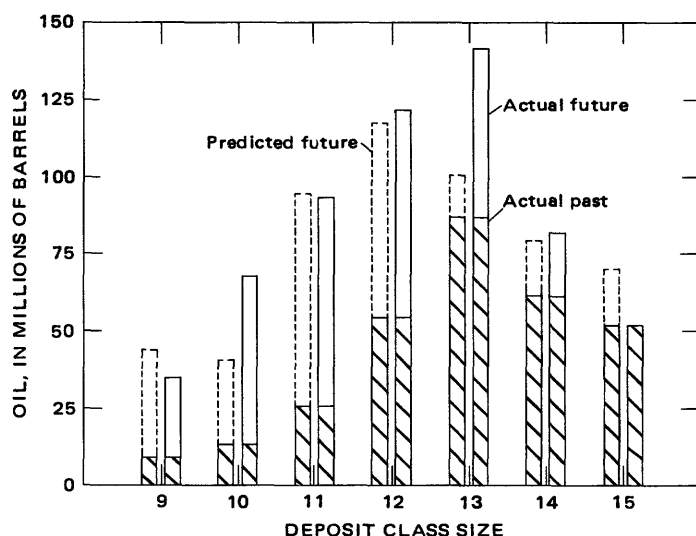


FIGURE 12.—Histogram comparing actual versus predicted petroleum discovered within size classes 9 through 15 for discoveries made between January 1, 1955, and December 31, 1969, based on data to December 31, 1954.

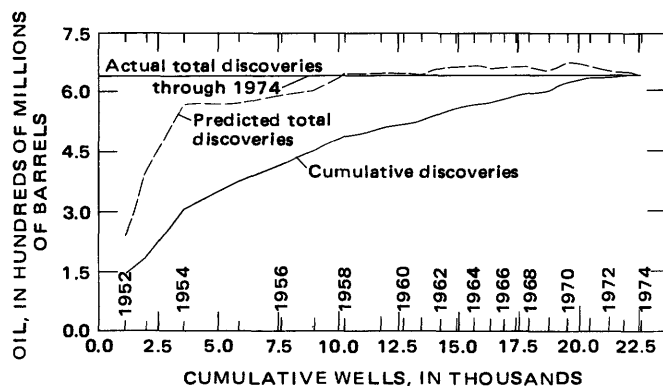


FIGURE 13.—Twenty-two successive predictions of the total volume of petroleum to have been discovered by the end of 1974 in classes 9 through 15.

DISCUSSION

In assessing the petroleum potential of an area as described above, no use was made of the geologic and geophysical knowledge acquired during the exploration of the area. Several reasons exist for not making direct use of such information. Geologic and geophysical data are difficult to assemble because they are scattered among many exploration companies and they are often confidential. Even if the data were gathered, a resource estimate for a partially explored area based upon them would have to be taken on faith because the data set would be too large and complex to communicate. Moreover, the reasoning by which one proceeds from geologic and geophysical knowledge of a basin to an estimate of the number and sizes of undiscovered

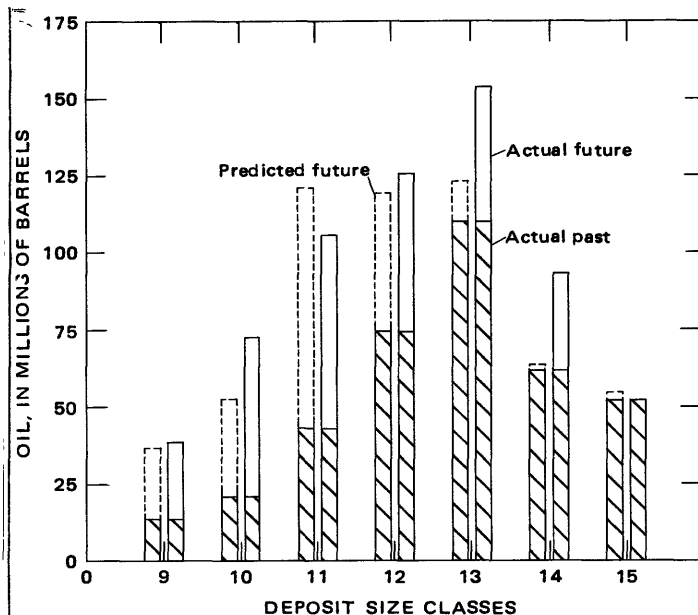


FIGURE 14.—Histogram comparing actual versus predicted petroleum discovered within size classes 9 through 15 for discoveries made between January 1, 1956, and December 31, 1974, based on data to December 31, 1955.

fields and the number of wells required to find them is largely subjective and intuitive and hence difficult to expose.

Although direct use of geologic and geophysical information is not feasible, this information is contained implicitly in the record of drilling and discovery insofar as it affected the siting of wells. For example, in the Denver basin, the effective basin size is determined by geologists and is revealed through the siting of wells. The effective basin area was estimated here to be about 10,000 mi², or about one-fifth of the whole Denver basin. This area was deduced without directly examining the reports on prospects in the other four-fifths of the basin. The examination of those reports was left to those who were siting the exploratory holes.

From the discovery rate, one can estimate how well exploration prospects are evaluated. The decline in the discovery rate (quantity of oil found per unit area searched) gives a quantitative measure of the superiority of actual exploration over random drilling. This superiority, measured by the parameter "efficiency of exploration," shows to what extent the better prospects are drilled early. The efficiency of exploration in the Denver basin was found to be about two to four times better than that of random drilling in the search for targets in larger classes (2×10^6 bbl to 32×10^6 bbl) and to be about the same as that of random drilling in the search for small targets. Thus, both the area to be explored and the efficiency with which it is being explored can be deduced from the drilling and discovery

record. When the area to be searched and the efficiency of search are known, one can extrapolate the past discovery record to obtain a reasonably accurate forecast of the number and sizes of oil fields that will be discovered by a given amount of future exploratory drilling.

For the Denver basin, the discovery record of the first 3,638 exploratory wells was extrapolated to provide forecasts of discoveries to be made by the next 7,929 exploratory wells. The future discoveries were divided into six different size classes. In most of these size classes, close agreement was observed between the predicted and actual discoveries. Estimates of future oil discoveries are difficult to make in the larger size classes because of the paucity of data. However, at the time in the discovery sequence when sufficient data are available to estimate future oil discoveries, most of the large deposits have been discovered, and the interest is in predicting the number of moderate size deposits remaining to be discovered.

REFERENCES CITED

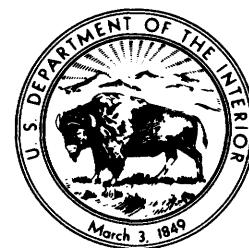
- Arps, J. J., and Roberts, T. G., 1958, Economics of drilling for Cretaceous oil and east flank of Denver-Julesburg basin: *American Association of Petroleum Geologists Bulletin*, v. 42, no. 11, p. 2549-2566.
- Oil and Gas Journal, 1969, Union Pacific, Pan Am enter exploration pact in Rockies: *Oil and Gas Journal* v. 67, no. 35, p. 86.
- Root, D. H., and Schuenemeyer, J., 1980, Petroleum-resource appraisal and discovery rate forecasting in partially explored regions—Mathematical foundations: U.S. Geological Survey Professional Paper 1138-B, p. B1-B9.
- Schuenemeyer, J. H., and Drew, L. J., 1977a, An exploratory drilling exhaustion plot program: *Computers and Geosciences*, v. 3, no. 4, p. 617-631.
- 1977b, Data reduction and analysis for a petroleum exploration model (abs.): American Association of Petroleum Geologists Conference, Exploration Data Synthesis, Tucson, March 1977, p. 1.
- Singer, D. A., 1976, Resin—A Fortran IV program for determining the area of influence of samples or drill holes in resource target search: *Computers and Geosciences*, v. 2, no. 2, p. 249-260.
- Singer, D. A., and Drew, L. J., 1976, The area of influence of an exploratory hole: *Economic Geology*, v. 71, no. 3, p. 642-647.

Petroleum-Resource Appraisal and Discovery Rate Forecasting in Partially Explored Regions— Mathematical Foundations

By D. H. ROOT *and* J. H. SCHUENEMEYER

G E O L O G I C A L S U R V E Y P R O F E S S I O N A L P A P E R I 1 3 8 - B

*The basic mathematical properties of a model
of the discovery process used to predict
the size distribution of future petroleum
deposits in a partially explored region*



CONTENTS

	Page
Abstract	B1
Introduction.....	1
Estimation of basin size	2
Practical difficulties and modifications	4
Efficiency of exploration	5
Results of simulation	7
References cited	9

ILLUSTRATION

FIGURE 1.	Estimates of effective basin size from simulated data.....	Page
		B8

TABLE

TABLE 1.	Estimated median efficiencies	Page
		B7

PETROLEUM-RESOURCE APPRAISAL AND DISCOVERY RATE FORECASTING IN PARTIALLY EXPLORED REGIONS— MATHEMATICAL FOUNDATIONS

By D. H. ROOT and J. H. SCHUENEMEYER

ABSTRACT

A stochastic model of the discovery process has been developed to predict, using past drilling and discovery data, the distribution of future petroleum deposits in partially explored basins, and the basic mathematical properties of the model have been established. The model has two exogenous parameters, the efficiency of exploration and the effective basin size. The first parameter is the ratio of the probability that an actual exploratory well will make a discovery to the probability that a randomly sited well will make a discovery. The second parameter, the effective basin size, is the area of that part of the basin in which drillers are willing to site wells. Methods for estimating these parameters from locations of past wells and from the sizes and locations of past discoveries were derived, and the properties of estimators of the parameters were studied by simulation.

INTRODUCTION

This paper presents the mathematical derivations for the estimators of the parameters used in a model of the discovery process proposed by Drew, Schuenemeyer, and Root (1980). The model was designed to summarize the past exploration and discovery experience in a region in order to predict the quantity of recoverable oil that future exploration could be expected to find. The model was developed because a significant amount of the petroleum remaining to be discovered in the United States and elsewhere is in regions where there has already been some exploratory drilling.

During exploration, a large amount of data is generated, including seismic surveys, magnetic surveys, geologic maps, well logs, discovery sizes and locations, and drilling dates and locations of discovery wells and dry holes. Because it is not feasible to use all the available data, a subset must be selected. The data incorporated in the model were the dates and locations of past exploratory wells together with the locations of discovery wells and the sizes of any deposits that were found. The geologic and geophysical data are thus implicitly included in that they guide the drilling and discovery process.

If exploration were completely random, then the probability of a given oil field being found by a single exploration well is the ratio of the area of the field to the area being searched. A model developed by Arps and Roberts (1958) assumes that the probability of an exploratory well discovering a given oil field is the product of a constant, the efficiency of exploration, and the ratio of the area of the field to the search area. In that model the measure of exploration was the number of exploratory wells. The model proposed by Drew, Schuenemeyer, and Root (1980) is similar to the model of Arps and Roberts but uses a different measure of past exploratory effort. Rather than the number of exploratory wells, it uses the cumulative area exhausted by these wells (Singer and Drew, 1976). The area exhausted by a well is related to the area and shape of the deposits under consideration. For example, for circular deposits having a radius of 1 mi, a dry exploratory well tells the explorer that the center of such a target cannot lie within 1 mi of the well and 3.14 mi^2 have been exhausted so far as such deposits are concerned. A pattern of dry wells will exhaust, with respect to such targets, an area equal to the area covered by all circles of radius 1 mi centered at each well. Because these circles may overlap, the exhausted area may be less than the number of wells times the area of a circle. If the deposits are elliptical having a major axis of, say, 4 mi and a minor axis of 2 mi, then a dry well exhausts all of the area within 1 mi of the well, none of the area more than 2 mi away, and partially exhausts the area between 1 and 2 mi from the well. A deposit centered within 1 mi from the well would have been hit regardless of its orientation, assuming certainty of recognition. If the center of the deposit is between 1 and 2 mi from the exploratory well, the deposit would have been hit for some orientations but not for others. The degree to which a point is exhausted is the probability that a randomly oriented deposit centered at the given point was hit. This concept of area exhausted can be used as a measure of the extent of exploration for targets of any given size and shape.

When an exploratory well is successful, then any point within 2 mi of the discovered deposit would have been partially exhausted for elliptical targets 4 mi by 2 mi. The degree of exhaustion at a point is the probability that a randomly oriented 4-mi by 2-mi ellipse centered at the point would have intersected the discovered deposit. This example illustrates that the area exhausted is a function of the size and shape of the deposits being considered and that more area is exhausted for larger deposits than for smaller ones. In the analysis of the discovery data, deposits are divided into size classes. All deposits within a given size class have approximately the same area, and each of these classes is analyzed independently. The following discussion focuses on elliptical deposits of the same size and shape.

If exploratory wells are randomly sited, then the petroleum richness of the area exhausted remains approximately equal to the richness of the unexplored area so that the fraction of all the deposits that are still undiscovered remains approximately equal to the fraction of the search area that has not been exhausted. If the exploration is better than random, then the petroleum richness of the area exhausted is greater than that of the area remaining to be searched.

In order to predict future discoveries, it is necessary to know the ultimate area of exhaustion that may be expected. This search area, defined as that part of the basin where operators are actually willing to site wells, is called the effective basin, and it may be only a small fraction of the entire basin. Because the size of the effective basin is an unknown quantity that typically grows through time, it was estimated from growth of the physical exhaustion time series.

ESTIMATION OF BASIN SIZE

We assume that within a large geologic basin there is a smaller effective basin to which the search for oil is restricted. Depth is not considered; the targets and the basin are assumed to be two-dimensional. The oil fields are assumed to be nonoverlapping ellipses all of the same size and shape inside the effective basin. The orientation of the targets, the angle between the major axis of the ellipse and the east-west line measured counter clockwise from the east-west line, may be anywhere between 0° and 180° . The search consists of the searcher selecting a point (drilling a well) in the effective basin. If the point is in a target, then the searcher is told the location of the boundaries of the target. This information approximates what would be learned in practice by subsequent development drilling.

There is an (x, y) coordinate system on the basin having x increase to the east and y increase to the north. The position of a target is completely specified by three coordinates (x, y, θ) , where x and y specify the location of the center and θ gives the angle between the major axis of the ellipse and the x -axis of the coordinate system measured counterclockwise from the x -axis. We define a search solid, S , to be all points (x, y, θ) , where (x, y) is a point in the geologic basin and θ is between 0° and 180° . Then any target in the geologic basin corresponds to a unique point in S and vice versa. We can visualize S as a cylinder 180 units high and having a base of the same size and shape as the geologic basin. Thus the θ -axis is vertical, and the x - y plane is horizontal.

When a search point (x', y') in the effective basin is selected and is not in a target, then every point (x, y, θ) in S that corresponds to an ellipse in the geologic basin that contains the search point, (x', y') , is eliminated from further consideration. The set of points in S condemned by an unsuccessful search point at (x', y') can be visualized as the volume swept out by a horizontal target-size ellipse centered at $(x', y', 0^\circ)$ with major axis lying along the x -axis as it rises to $(x', y', 180^\circ)$ while rotating 1° , counterclockwise from above, for each degree it rises. Each point in this volume corresponds to an ellipse in the geologic basin that would contain (x', y') . When a search point is successful, then the points in S that are eliminated are all those points corresponding to ellipses that would intersect the target which was found. The introduction of the space S has changed the problem from search points looking for area targets to search volumes in S looking for point targets in S . The cumulative volume in S that has been eliminated by successful and unsuccessful search points in the effective basin, EB , is measured in acre-degrees. This unit, the acre-degree, is the unit in which we will measure the extent of exploration.

Inside the search volume, S , we define the effective search volume, ES , as all points (x, y, θ) in S where (x, y) is a point in the effective basin. The size of the effective basin is not known to the data analyst. We will give a method for estimating the effective basin size from well locations under simplifying assumptions and then describe how the method is modified to fit more realistic assumptions. At first, it is assumed that wells are sited randomly in EB according to a uniform probability distribution. From this assumption we derive an estimator of effective basin size which is then modified to account for the fact that development wells are always close to existing wells.

Let EB be the effective basin, let β be the Lebesgue measurable sets on EB , and assume EB is endowed with the uniform probability measure. Let $\Omega =$

$EB \times EB \times \dots$, a countable product of copies of the effective basin. Let β_m be the smallest σ field containing all sets of the form $A_1 \times \dots \times A_m \times EB \times \dots$ where each A_i is a Lebesgue measurable set in EB . Let β_0 be the trivial σ field on Ω . A function defined on Ω can be β_m measurable only if it is dependent on at most the first m coordinates of a point of Ω . Let $d_j, j = 1, 2, \dots$ be a sequence of positive numbers. Let $D_j(P)$ denote a disk of area d_j centered at the point P . If (P_1, P_2, \dots) is a point in Ω , define $a(j, n; P_1, P_2, \dots)$ to be the total area covered by $D_j(P_1), \dots, D_j(P_n)$. Because $a(j, n)$ is a continuous function of P_1, \dots, P_n it is measurable with respect to β_n .

Using the notation just defined we now establish the basic theorem for the estimation of effective basin size. **Theorem:** If $d_j \rightarrow 0$ as $j \rightarrow \infty$ and n and j are made to approach infinity in such a way that $nd_j \rightarrow A$ where A is an arbitrarily selected positive number and if EB is a bounded set whose boundary has zero area, then

$$\frac{a(j, n)}{EB(1 - e^{-A/EB})} \rightarrow 1 \text{ in probability.}$$

The symbol EB is used for both the effective basin and its area.

Proof: The proof proceeds by first showing that $E(a(j, n)) \rightarrow EB(1 - e^{-A/EB})$ and then showing that $\sigma^2(a(j, n)) \rightarrow 0$. Let a sequence $r_j, j = 1, 2, \dots$ of positive numbers be defined by the equation $\pi r_j^2 = d_j$. Define $EB(r_j +)$ to be all points (x, y) for which there is a point (x', y') in EB such that $(x' - x)^2 + (y' - y)^2 \leq r_j^2$. Define $EB(r_j -)$ to be all points (x, y) in EB for which there is no point (x', y') outside of EB such that $(x' - x)^2 + (y' - y)^2 \leq r_j^2$.

Thus $EB(r_j +)$ is EB swelled by r_j and $EB(r_j -)$ is EB shrunk by r_j . Define $\partial EB(r_j)$ by $\partial EB(r_j) = EB(r_j +) - EB(r_j -)$. Thus $\partial EB(r_j)$ is all points within r_j of the boundary of EB . Let f_j be a function defined over $EB(r_j +)$ by

$$f_j(P) = \begin{cases} d_j/EB & \text{for } P \text{ in } EB(r_j -) \\ \text{area}(D_j(P) \cap EB)/EB & \text{for } P \text{ in } \partial EB(r_j). \end{cases}$$

Then $f_j(P)$ is the probability that a randomly located disk D_j will cover the point P . Let $S(j, n; P_1, P_2, \dots)$ be the set in $EB(r_j +)$ actually covered by the first n disks $D_j(P_1), \dots, D_j(P_n)$. We now establish an inequality by successive conditioning. From the definition of conditional expectation we have, letting $S = S(j, n-1)$,

$$E(a(j, n) | \beta_{n-1}) = d_j + a(j, n-1) - \int_S f_j(P) dP; \quad (1)$$

since $f_j \leq d_j/EB$ we have

$$\int_S f_j(P) dP \leq a(j, n-1) d_j/EB.$$

Substituting this expression for the integral yields

$$E(a(j, n) | \beta_{n-1}) \geq d_j + (1 - d_j/EB) a(j, n-1). \quad (2)$$

Next we condition both sides with respect to β_{n-2} ; recall that

$$E(E(a(j, n) | \beta_{n-1}) | \beta_{n-2}) = E(a(j, n) | \beta_{n-2}) \geq d_j + (1 - d_j/EB) E(a(j, n-1) | \beta_{n-2}). \quad (3)$$

Applying our inequality (1) to $E(a(j, n-1) | \beta_{n-2})$ gives

$$E(a(j, n-1) | \beta_{n-2}) \geq d_j + (1 - d_j/EB) a(j, n-2). \quad (4)$$

Combining (3) and (4) gives

$$E(a(j, n) | \beta_{n-2}) \geq d_j + (1 - d_j/EB)(d_j + (1 - d_j/EB) a(j, n-2)) = d_j + (1 - d_j/EB) d_j + (1 - d_j/EB)^2 a(j, n-2).$$

Continuing this process through $n-2$ more conditionings gives

$$E(a(j, n)) = E(a(j, n) | \beta_0) \geq d_j \sum_{k=0}^{n-2} (1 - d_j/EB)^k \quad (5)$$

and

$$d_j \sum_{k=0}^{n-2} (1 - d_j/EB)^k = EB - EB(1 - d_j/EB)^{n-1}. \quad (6)$$

Because we assume that $nd_j \rightarrow A$ we have that the R.H.S. of (6) approaches $EB - EB e^{-A/EB}$. Thus we may conclude

$$\lim_{n, j} E(a(j, n)) \geq EB(1 - e^{-A/EB}) \quad (7)$$

when $nd_j \rightarrow A$ as $n, j \rightarrow \infty$

Referring back to equation 1, we now seek a lower bound for $\int_S f_j(P) dP$.

Recall that f_j is d_j/EB on $EB(r_j -)$ and it takes smaller values on $\partial EB(r_j)$ so we assume that as much of $S(j, n-1)$ is in $\partial EB(r_j)$ as possible to find a lower bound. Thus

$$\int_S f_j(P) dP = \int_S f_j(P) dP + \int_{S \cap \partial EB(r_j)} f_j(P) dP \geq \int_{S \cap EB(r_j -)} f_j(P) dP$$

The area of $S(j, n-1) \cap EB(r_j -)$ is at least $a(j, n-1) - \partial EB(r_j)$, where $\partial EB(r_j)$ is used to denote the area of $\partial EB(r_j)$. Thus $\int_S f_j(P) dP \geq (a(j, n-1) - \partial EB(r_j)) d_j/EB$,

which yields the inequality,

$$\begin{aligned}
E(a(j,n) | \beta_{n-1}) &\leq \\
d_j + a(j,n-1) - (a(j,n-1) - \partial EB(r_j))d_j/EB \\
&= d_j + a(j,n-1)(1-d_j/EB) + (\partial EB(r_j))d_j/EB.
\end{aligned}$$

As before we condition both sides on β_{n-2} and obtain

$$\begin{aligned}
E(a(j,n) | \beta_{n-2}) &\leq d_j + (1-d_j/EB)E(a(j,n-1) | \beta_{n-2}) + \\
&(\partial EB(r_j))d_j/EB \leq d_j + (1-d_j/EB)(d_j + (1-d_j/EB) \times \\
&a(j,n-2) + (\partial EB(r_j))d_j/EB) + (\partial EB(r_j))d_j/EB \\
&= d_j + (1-d_j/EB)d_j + (\partial EB(r_j))d_j/EB \\
&+ (1-d_j/EB)^2 a(j,n-2) \\
&+ (1-d_j/EB) (\partial EB(r_j))d_j/EB.
\end{aligned}$$

We repeat this process for $\beta_{n-3}, \beta_{n-4}, \dots, \beta_0$ to obtain

$$\begin{aligned}
E(a(j,n)) &\leq d_j \sum_{k=0}^{n-1} (1-d_j/EB)^k \\
&+ (\partial EB(r_j))d_j/EB \times \sum_{k=0}^{n-1} (1-d_j/EB)^k. \quad (8)
\end{aligned}$$

Since the boundary of EB has 0 area and since EB is a bounded set it follows that $\partial EB(r_j) \rightarrow 0$ as $r_j \rightarrow 0$. Hence as $n, j \rightarrow \infty$ the second term in (8) goes to 0. The limit of the first term was calculated before (see equation (6)) to be $EB(1-e^{-A/EB})$.

Hence we have shown that

$$\lim_{n,j} E(a(j,n)) \leq EB(1-e^{-A/EB}) \quad (9)$$

when $nd_j \rightarrow A$.

Combining (9) and (7) we have that

$$E(a(j,n)) \rightarrow EB(1-e^{-A/EB}) \text{ when } n, j \rightarrow \infty \text{ and } nd_j \rightarrow A.$$

To complete the proof of the theorem, it is sufficient to show that the variance of $a(j,n)$, $\sigma^2(a(j,n)) \rightarrow 0$ as $n, j \rightarrow \infty$ and $nd_j \rightarrow A$. This will be done by establishing

$$\sigma^2(a(j,n)) \leq nd_j^2. \quad (10)$$

By definition of variance

$$\sigma^2(a(j,n)) = E((a(j,n) - E(a(j,n)))^2).$$

We decompose the function $a(j,n) - E(a(j,n))$ into a sum of orthogonal functions

$$\begin{aligned}
a(j,n) - E(a(j,n)) &= \sum_{k=1}^n [E(a(j,n) | \beta_k) - E(a(j,n) | \beta_{k-1})].
\end{aligned}$$

Because the n summands are orthogonal, we have that

$$\begin{aligned}
E((a(j,n) - E(a(j,n)))^2) &= \sum_{k=1}^n [E((E(a(j,n) | \beta_k) - E(a(j,n) | \beta_{k-1}))^2)].
\end{aligned}$$

We will have completed the proof of (10), and hence of the theorem, when we have shown that

$$|E(a(j,n) | \beta_k) - E(a(j,n) | \beta_{k-1})| \leq d_j.$$

Let $(P_1, \dots, P_{k-1}, P_k, P_{k+1}, \dots, P_n, \dots) = P$ and $(P_1, \dots, P_{k-1}, P'_k, P_{k+1}, \dots, P_n, \dots) = P'$ be two points

in Ω which agree in each of the first n coordinates except for the k^{th} . Then $|a(j,n)(P) - a(j,n)(P')| \leq d_j$ because they differ only in the position of a single disk of area d_j . Hence

$$\begin{aligned}
|E(a(j,n) | \beta_k)(P_1, \dots, P_{k-1}, P'_k) - \\
E(a(j,n) | \beta_k)(P_1, \dots, P_k)| \leq d_j.
\end{aligned}$$

Hence it follows that

$$\begin{aligned}
|E(a(j,n) | \beta_{k-1})(P_1, \dots, P_{k-1}) - \\
E(a(j,n) | \beta_k)(P_1, \dots, P_k)| \leq d_j.
\end{aligned}$$

This last inequality completes the proof of (10). The theorem then follows from Chebyshev's inequality. End of proof.

Corollary: If, in the above theorem, the sets D_j are replaced by ellipses of area d_j all having the same orientation, then the conclusions of the theorem still hold if the major axis of $D_j \rightarrow 0$ as $j \rightarrow \infty$.

Proof: The only modification required in the theorem is to let r_j be equal to one-half of the major axis of D_j .

We can derive an estimator for EB from the above theorem and corollary. When n points in EB have been selected at random, disks of area d are centered at each point and the total area covered, $a(n)$, is measured. The estimate, \hat{EB} , for EB is then chosen to satisfy the relation

$$a(n) = \hat{EB} (1 - e^{-nd/\hat{EB}}). \quad (11)$$

PRACTICAL DIFFICULTIES AND MODIFICATIONS

There is the practical problem of what size disk to use. If d is too small, then $a(n) = nd$ from which it follows that $\hat{EB} = \infty$. On the other hand, if the disks are too large, then because $\hat{EB} > a(n) > d$, we will get unreasonably large estimates for the effective basin size. The problem of disk size is further complicated by

the effect of target size on the effective basin size. For example, a very large area may be sparsely drilled for very large targets, whereas the search for small targets might be restricted to a smaller area. The assumption of random location of wells is not applicable to development wells, which are always sited near existing wells, so the disks around development wells cover very little additional area compared to what the average randomly located disk would cover.

In applying the estimator, \hat{EB} , defined by equation (11) in the analysis of the discovery of targets of a particular size class, d was taken to be the area of the targets in question, $a(n)$ was taken to be the acre-degrees in the search solid which had been exhausted divided by 180, and n in the exponent was taken to be the number of wildcat wells plus a fraction, α , of the number of development wells. The fraction, α , is chosen to satisfy the condition that if D_1 and D_2 are target-size and target-shape ellipses having the same but randomly selected orientation and centers 0.25 mi apart (40-acre spacing), then $(1 + \alpha)d$ is the expected total area covered by D_1 and D_2 . Thus α is between 0 and 1 and decreases as target size increases.

EFFICIENCY OF EXPLORATION*

We assume that there are N target ellipses all of the same shape and area in EB and that they correspond to N points in ES , the effective search solid. Let T_i be the fraction of ES that had been exhausted at the time of the i^{th} discovery. Thus $0 < T_1 < T_2 < \dots < T_N < 1$. Because the search process is random, the T_i are random. We wish to find the joint probability distribution of these N random variables as a function of the efficiency of exploration. We say that exploration is c times as efficient as random if an infinitesimal amount of search is c times as likely to make a discovery as if it were purely random. Specifically, if " a " of ES has been searched and $N - k + 1$ targets remain to be found, then as the fraction of ES searched increases from " a " to " $a + \Delta a$," the probability of discovering exactly " j " targets in the case of a purely random search is

$$\binom{N-k+1}{j} \left(\frac{\Delta a}{1-a} \right)^j \left(\frac{1-a-\Delta a}{1-a} \right)^{N-k+1-j}.$$

From this expression we now find the density of T_k in the case of purely random search. Let $f_k(a | T_{k-1} = a_{k-1})$ be the conditional density of T_k , given that the $k-1$ discovery occurred at a_{k-1} . Clearly $f_k(a | T_{k-1} = a_{k-1}) = 0$ for $a < a_{k-1}$ or $a > 1$.

Suppose $a_{k-1} < a < 1$; then

$$\begin{aligned} f_k(a | T_{k-1} = a_{k-1}) &= \lim_{\Delta a \rightarrow 0} \frac{1}{\Delta a} P[a < T_k < a + \Delta a | T_{k-1} = a_{k-1}] \\ &= \lim_{\Delta a \rightarrow 0} \frac{1}{\Delta a} P[T_k < a + \Delta a | T_k > a \text{ and } T_{k-1} = a_{k-1}] \times P[T_k > a | T_{k-1} = a_{k-1}] \\ &= \lim_{\Delta a \rightarrow 0} \frac{1}{\Delta a} \left[\sum_{j=1}^{N-k+1} \binom{N-k+1}{j} \left(\frac{\Delta a}{1-a} \right)^j \left(\frac{1-a-\Delta a}{1-a} \right)^{N-k+1-j} \right] \times \left(\frac{1-a}{1-a_{k-1}} \right)^{N-k+1} \\ &= \frac{N-k+1}{1-a} \left(\frac{1-a}{1-a_{k-1}} \right)^{N-k+1} \end{aligned}$$

We isolate the factor

$$\lim_{\Delta a \rightarrow 0} \frac{1}{\Delta a} P[T_k < a + \Delta a | T_k > a \text{ and } T_{k-1} = a_{k-1}]$$

and call it the instantaneous rate of discovery for the k^{th} discovery, given that the next discovery is the k^{th} discovery. In purely random exploration, this instantaneous rate is

$$\frac{N-k+1}{1-a}.$$

We say that exploration is c times as efficient as random exploration if the instantaneous discovery rate for the k^{th} discovery is

$$c \left(\frac{N-k+1}{1-a} \right).$$

The instantaneous discovery rate for the k^{th} discovery determines the conditional density of T_k given T_{k-1} . For if $g_k(a | T_{k-1})$ is the conditional density for T_k given T_{k-1} and $h_k(a | T_{k-1})$ is the instantaneous discovery rate, then h_k and g_k must satisfy the relation

$$h_k(a | T_{k-1}) = \frac{g_k(a | T_{k-1})}{1 - \int_{T_{k-1}}^a g_k(s | T_{k-1}) ds}$$

whenever $a > T_{k-1}$. Thus h_k and g_k each determine the other. Using the above notation we now establish the

*A part of the results in this and the next section appeared in Schuenemeyer and Root (1977) and are reproduced here with permission from American Statistical Association.

main theorem from which the estimator of the efficiency is derived.

Theorem: If the efficiency is $c > 0$, then T_1, \dots, T_N have the same joint density as the order statistics of N independent random variables X_1, \dots, X_N taking values on the interval $[0, 1]$ and having the density $f(x) = c(1-x)^{c-1}$.

Moreover if $T_k < a < T_{k+1}$ for some fixed "a," then

$$-\sum_{i=1}^k \ln(1-T_i)$$

has the same density as the sum of k independent random variables Z_1, \dots, Z_k each having the density

$$h(z) = \begin{cases} \frac{ce^{-cz}}{1-(1-a)^c} & 0 < z < -\ln(1-a). \\ 0 & \text{otherwise} \end{cases}$$

Proof: Let X_1, \dots, X_N be independent random variables with common density

$$f(x) = \begin{cases} c(1-x)^{c-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

and distribution function $F(x) = 1-(1-x)^c$.

Let Y_1, Y_2, \dots, Y_N be their order statistics, that is, Y_1 is the minimum of X_1, X_2, \dots, X_N and Y_2 is the next larger and so forth, Y_N being the maximum of X_1, \dots, X_N .

In order to show that Y_1, \dots, Y_N and T_1, \dots, T_N have the same joint density, it is sufficient to show that

$$\lim_{\Delta a \rightarrow 0} \frac{1}{\Delta a} P[Y_k < a + \Delta a \mid Y_k > a \text{ and } Y_{k-1} < a] = \frac{c(N-k+1)}{1-a},$$

the instantaneous discovery rate for the k^{th} discovery.

$$\begin{aligned} & \lim_{\Delta a \rightarrow 0} \frac{1}{\Delta a} P[Y_k < a + \Delta a \mid Y_k > a \text{ and } Y_{k-1} < a] \\ &= \lim_{\Delta a \rightarrow 0} \frac{1}{\Delta a} \frac{P[Y_k < a + \Delta a \text{ and } Y_k > a \text{ and } Y_{k-1} < a]}{P[Y_k > a \text{ and } Y_{k-1} < a]} \\ &= \lim_{\Delta a \rightarrow 0} \frac{1}{\Delta a} \left(\frac{N}{k-1} \right) (F(a))^{k-1} \sum_{j=1}^{N-k+1} \binom{N-k+1}{j} \times \\ & \quad (F(a+\Delta a) - F(a))^j (1-F(a+\Delta a))^{N-k+1-j} \div \\ & \quad \left(\frac{N}{k-1} \right) (F(a))^{k-1} (1-F(a))^{N-k+1} \end{aligned}$$

$$= \lim_{\Delta a \rightarrow 0} \frac{1}{\Delta a} \frac{(N-k+1) (F(a+\Delta a) - F(a)) (1-F(a+\Delta a))^{N-k}}{(1-F(a))^{N-k+1}}$$

$$= \frac{(N-k+1) f(a) (1-F(a))^{N-k}}{(1-F(a))^{N-k+1}}$$

$$= \frac{(N-k+1) f(a)}{1-F(a)} = \frac{(N-k+1) c(1-a)^{c-1}}{(1-a)^c} = \frac{c(N-k+1)}{1-a}$$

Hence Y_1, \dots, Y_N and T_1, \dots, T_N have the same joint density. The last part of the theorem follows from the fact that Z_i has the density of $-\ln(1-X_i)$ given that $X_i < a$ and $-\sum \ln(1-T_i)$ has the same probability

distribution as $-\sum \ln(1-Y_i)$ which has the same probability distribution as $-\sum \ln(1-X_i)$.

The joint density of T_1, \dots, T_N is given by

$$f(a_1, \dots, a_N) = N! \ c^N \prod_{i=1}^N (1-a_i)^{c-1}$$

The conditional density of T_1, \dots, T_k , given that there are exactly k discoveries in the exploration of "a" of the search solid, is given by

$$\frac{\int_a^{a+\Delta a} \dots \int_a^{a+\Delta a} f(a_1, \dots, a_N) da_{k+1} \dots da_N}{P[T_k < a \text{ and } T_{k+1} > a]} = \frac{k! \ c^k \prod_{i=1}^k (1-a_i)^{c-1}}{(1-(1-a)^c)^k}$$

Two important features of this joint density are that it is independent of N and that a sufficient statistic for "c" is given by

$$\prod_{i=1}^k (1-a_i).$$

Hence $-\sum_{i=1}^k \ln(1-a_i)$ is also a sufficient statistic

for c given that there are k discoveries. From the previous theorem we can calculate that

$$E \left[-\frac{1}{k} \sum_{i=1}^k \ln(1-t_i) \right] = \frac{1}{k} \sum_{i=1}^k E(Z_i) \\ = \frac{1}{c} + \frac{(1-a)^c \ln(1-a)}{1-(1-a)^c}$$

We can thus obtain a point estimate, \hat{c} , for c based on a_1, \dots, a_k by choosing \hat{c} so that

$$-\frac{1}{k} \sum_{i=1}^k \ln(1-a_i) = \frac{1}{\hat{c}} + \frac{(1-a)^{\hat{c}} \ln(1-a)}{1-(1-a)^{\hat{c}}} \quad (12)$$

Such a \hat{c} will exist only if

$$-\frac{1}{k} \sum_{i=1}^k \ln(1-a_i) < \frac{-\ln(1-a)}{2},$$

and this inequality need not always hold.

In the practical problem of oil exploration (see Drew and others, 1980) it is reasonable to assume that $c \geq 1$. Therefore, in the analysis of oil exploration data we take \hat{c} to be the solution of (12) if (12) has a solution and if that solution is greater than or equal to 1; otherwise \hat{c} is taken to be 1.

The unconditional joint density of the areas exhausted by the time that "a" of the effective search solid has been exhausted is

$$h(a_1, \dots, a_k) = \frac{k! c^k \prod_{i=1}^j (1-a_i)^{c-1}}{(1-(1-a)^c)^k} \\ \times \binom{N}{k} (1-(1-a)^c)^k ((1-a)^c)^{N-k},$$

and its integral over all (a_1, \dots, a_k) is the probability of exactly k discoveries. If one attempts to find maximum likelihood estimates simultaneously for "N" and "c," the maximum can occur at $N = +\infty$ and $c = 0$. To avoid this, it is necessary to include in the model the condition that the point targets in the effective search solid cannot be too close together because they must correspond to nonoverlapping ellipses in the effective basin. Because this condition is difficult to work with, it is decided to estimate the efficiency first by the method described above and then to choose \hat{N} to maximize $\frac{\hat{N}}{k}$

$(1-(1-a)^c)^k ((1-a)^c)^{\hat{N}-k}$, which means \hat{N} is the smallest integer such that

$$\hat{N} > \frac{k}{1-(1-a)^{\hat{c}}} - 1.$$

RESULTS OF SIMULATION

A simulation study was conducted to investigate the behavior of the estimated effective basin size. The well locations were chosen randomly within the basin, and the basin size was estimated using equation (11). Circular disks were assumed; the ratio of the area of a single target to the basin size was 1 to 2,500. The results for three replications are presented in figure 1; basin size estimates are plotted for every tenth well. The estimated basin size is infinite until disks overlap.

In order to study the distribution of \hat{c} as estimated by equation (12), a simulation experiment was conducted for various values of c and N . Estimates for c were calculated at 0.1, 0.2, \dots , 0.9 fraction of the area searched, given that at least one target had been discovered. The median efficiencies from these simulations are presented in table 1. Each simulation was replicated 100 times; however, the median estimates corresponding to $c=2$ or $c=4$ sometimes are based on fewer replications when the area searched is less than 0.4 since at least one discovery is required for an estimate. The estimated efficiency was set to zero when discoveries were present, but no solution to (12)

TABLE 1.—Estimated median efficiencies

Efficiency = 1.0			
Fraction of area searched	Medians $N = 50$	Medians $N = 70$	Medians $N = 90$
0.1	0.33	0.00	0.82
.2	.88	.45	1.07
.3	1.17	1.12	.92
.4	.83	.94	1.08
.5	.97	.96	1.04
.6	.99	1.13	1.04
.7	.94	1.06	1.00
.8	.92	1.03	1.10
.9	1.01	1.01	1.03
Efficiency = 2.0			
Fraction of area searched	Medians $N = 20$	Medians $N = 30$	Medians $N = 40$
0.1	1.27	2.30	0.00
.2	1.33	1.83	1.95
.3	2.70	2.36	1.81
.4	2.08	2.17	2.16
.5	1.95	2.11	2.20
.6	2.02	1.91	2.02
.7	2.17	2.01	1.98
.8	2.16	1.98	2.02
.9	2.05	1.97	2.03
Efficiency = 4.0			
Fraction of area searched	Medians $N = 3$	Medians $N = 5$	Medians $N = 10$
0.1	5.30	6.18	3.22
.2	3.31	3.64	4.73
.3	4.43	4.21	4.29
.4	3.80	4.39	4.47
.5	4.02	3.84	4.32
.6	4.25	4.21	4.53
.7	4.19	4.41	4.52
.8	4.31	4.35	4.55
.9	4.33	4.37	4.54

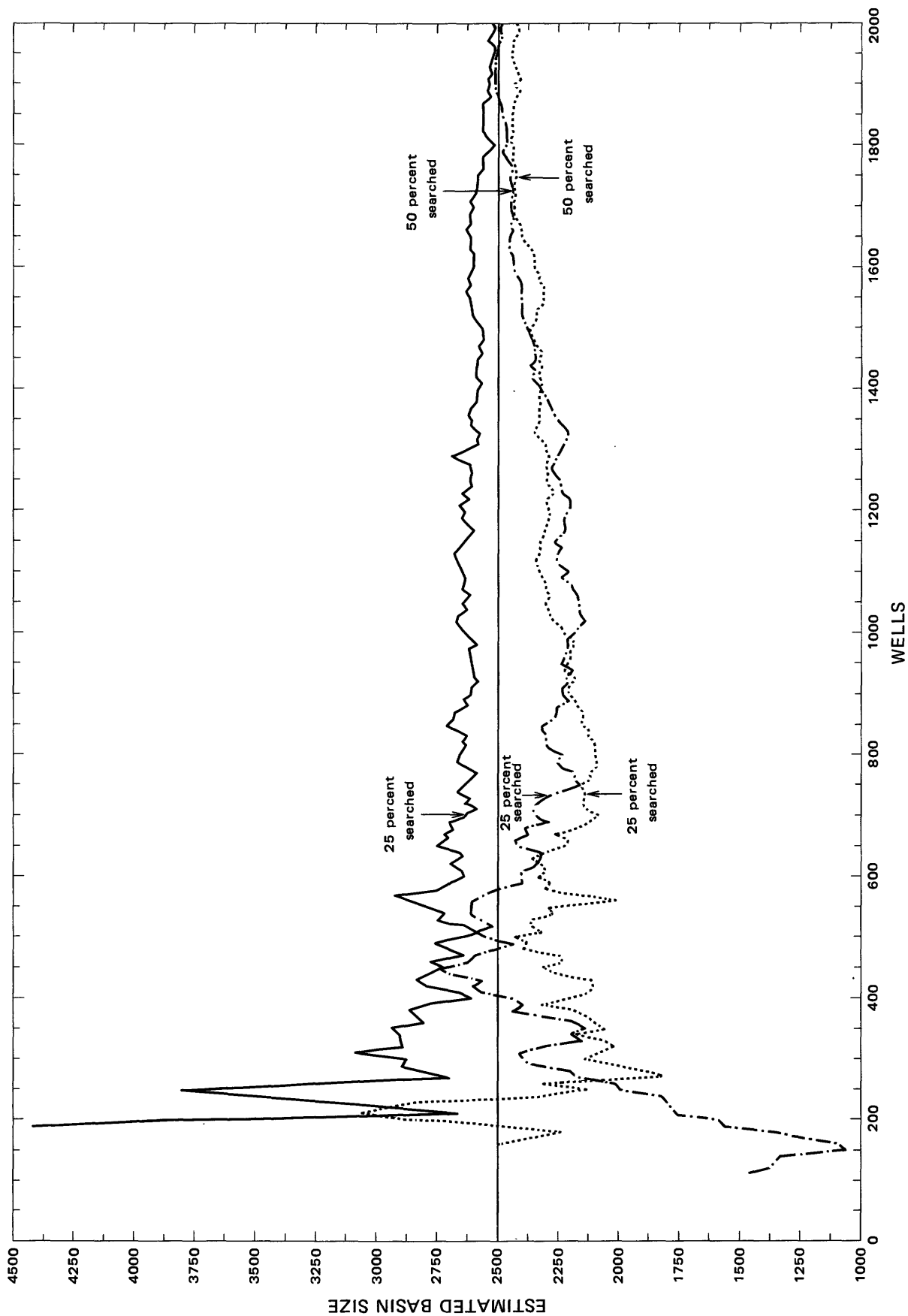


FIGURE 1.— Three basin size simulations showing estimated basin size plotted against the number of wells on which the estimate was based. The actual basin size was 2,500 and the disk size was 1.0.

existed. The fractions of the area searched were generated from $1-(1-u_i)^{1/c}$ where u_i is uniform (0,1), $i=1,2,\dots,N$. Smaller values of N were associated with larger c 's because in petroleum exploration the larger, less numerous deposits tend to be found with higher efficiency.

REFERENCES CITED

- Arps, J. J., and Roberts, T. G., 1958, Economics of drilling for Cretaceous oil and east flank of Denver—Julesburg basin: American Association of Petroleum Geologists Bulletin, v. 42, no. 11, p. 2549–2566.
- Drew, L. J., Schuenemeyer, J. H., and Root, D. H., 1980, Petroleum-resource appraisal and discovery rate forecasting in partially explored regions—An application to the Denver basin: U.S. Geological Survey Professional Paper 1138–A, p. A1–A11.
- Schuenemeyer, J. H., and Root, D. H., 1977, Computational aspects of a probabilistic oil discovery model: Proceedings of the Statistical Computing Section, American Statistical Association, p. 347–351.
- Singer, D. A., and Drew, L. J., 1976, The area of influence of an exploratory hole: Economic Geology, v. 71, no. 3, p. 642–647.

Petroleum-Resource Appraisal and Discovery Rate Forecasting in Partially Explored Regions —

An Application to Supply Modeling

By E. D. ATTANASI, L. J. DREW, and J. H. SCHUENEMEYER

GEOLOGICAL SURVEY PROFESSIONAL PAPER 1138-C

*Investigation of field exploration behavior
with distributed lag models and a discovery
process model, and application of models to
predict forthcoming reserves*



UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1980

CONTENTS

	Page
Abstract	C1
Introduction	1
Previous studies	2
Operator behavior at the field level	3
Firm behavior	4
Formation of expectations	4
Distributed lag models	5
Operator drilling behavior	7
Empirical results	8
Application of the discovery process model predictions	11
Supply of reserves	15
Conclusions	18
References cited	19

ILLUSTRATIONS

FIGURE		Page
1.	Shapes of normalized time profiles for weights associated with values of the lag variable	C7
2.	Temporal effects of a one-time unit change in the value of total discoveries on wildcat drilling	11
3.	Temporal effects of a one-time unit change in the value of total discoveries on drilling expenditures	11
4.	Residuals for well-drilling models	14
5.	Residuals for well-drilling expenditures models	15

TABLES

TABLE		Page
1.	Estimated coefficients of infinite lag models	C9
2.	Coefficients of polynomial distributed lag models semiannual observations	10
3.	Coefficients of polynomial distributed lag models quarterly data	10
4.	Forecast performance of integrated system	16

PETROLEUM-RESOURCE APPRAISAL AND DISCOVERY RATE FORECASTING IN PARTIALLY EXPLORED REGIONS— AN APPLICATION TO SUPPLY MODELING

By E. D. ATTANASI, L. J. DREW, and J. H. SCHUENEMEYER

ABSTRACT

This study examines the temporal properties and determinants of petroleum exploration for firms operating in the Denver basin. Expectations associated with the favorability of a specific area are modeled by using distributed lag proxy variables (of previous discoveries) and predictions from a discovery process model. In the second part of the study, a discovery process model is linked with a behavioral well-drilling model in order to predict the supply of new reserves.

Results of the study indicate that the positive effects of new discoveries on drilling increase for several periods and then diminish to zero within 2½ years after the deposit discovery date. Tests of alternative specifications of the argument of the distributed lag function using alternative minimum size classes of deposits produced little change in the model's explanatory power. This result suggests that, once an exploration play is underway, favorable operator expectations are sustained by the quantity of oil found per time period rather than by the discovery of specific size deposits. When predictions of the value of undiscovered deposits (generated from a discovery process model) were substituted for the expectations variable in models used to explain exploration effort, operator behavior was found to be consistent with these predictions. This result suggests that operators, on the average, were efficiently using information contained in the discovery history of the basin in carrying out their exploration plans. Comparison of the two approaches to modeling unobservable operator expectations indicates that the two models produced very similar results. The integration of the behavioral well-drilling model and discovery process model to predict the additions to reserves per unit time was successful only when the quarterly predictions were aggregated to annual values. The accuracy of the aggregated predictions was also found to be reasonably robust to errors in predictions from the behavioral well-drilling equation.

INTRODUCTION

Interest in the state of the U.S. domestic petroleum industry has resulted in a critical examination of the usefulness of current economic models. There is some question as to whether these models can provide useful information about petroleum supply availability, supply price sensitivity, and the effects of alternative policy options on future supply possibilities. The performance of current models in terms of predicting future oil and gas discoveries and supply has thus far been disappointing (MacAvoy and Pindyck, 1975). The poor performance of these models may be caused in

part by their almost universal reliance on aggregated data, which obscures the effects of physical exhaustion on the supply response.

Although a major reason for constructing empirical economic models is to predict price responsiveness of future supply, the historical data upon which current models are based generally do not contain sufficient price variation to accurately predict future price-supply responses within a region. The price variations present in available data relate to the quality of crude oil rather than to the incremental costs of exploration and development of individual deposits. Moreover, the effect of price changes on expected supplies is moderated by the level of resource depletion for a particular basin. Because of the potential interaction between price and resource depletion, the use of highly aggregated data will likely lead to spurious correlation between changes in supply and incremental price changes. Very few attempts have been made to model economic behavior at the level of the exploration play. In this report, an exploration play is defined as the increase in wildcat drilling attributed to and following the discovery of a significant (large) deposit in a formation that was not known to yield significant amounts of oil. Although the major obstacle to modeling has been the lack of basin-specific exploration data, many workers believe that field behavior is too erratic or unsystematic to model successfully.

This study has two parts. The first discusses the temporal properties of operators' exploration behavior for the Denver basin. In particular, the field behavior of operators in terms of responses (wildcat drilling and drilling expenditures) to new discoveries is examined and then compared with results from another basin. Properties of operators' behavior include the duration of the response and the nature of new deposits that must be discovered in order to sustain favorable operator expectations associated with a given area. The stochastic process characterizing the temporal distribution of previous discoveries is used as a basis for modeling operator expectations associated with the

value of deposits remaining to be found. The stochastic nature of this scheme is compared with the process characterizing the predictions of the value of remaining deposits calculated with discovery process models similar to the ones proposed by Arps and Roberts (1958) and Drew, Schuenemeyer, and Root (1980).

In the second part of the study, the discovery process model is linked with the drilling model in a recursive system of equations in order to predict the supply of new reserves. Whereas the purpose of the behavioral drilling model is to explain operator action within a given time period, the discovery process model predicts the number and size distribution of new discoveries as a function of the number of wells drilled. By linking the two models, we can predict the quantity of new reserves discovered per unit time. These predictions are compared to the actual sequence of discoveries.

PREVIOUS STUDIES

Industry-level petroleum supply models have historically treated exploration as a process using wildcat wells that generates proved reserves. As such, little effort was devoted to formulating testable behavioral hypotheses or extending the theory of the firm to include the exploration function. In early models, no consideration was given to integrating the role of the initial resource base into the analysis. Because of geologic differences from area to area and very limited historical data series, the results of those industry-level studies have been somewhat disappointing. Data have been aggregated to the extent that significant behavioral and physical relations are concealed by the model specifications. Moreover, although almost all of these models have attempted to determine the price sensitivity of exploration and petroleum supply, price has varied little within specific geologic regions during the time periods chosen for analysis. Alternatively, engineering-type process models (National Petroleum Council, 1972; Federal Energy Administration, 1974; 1976) designed to forecast oil and gas supply have frequently reduced exploration behavior to mechanical rules, assuming, for example, that firms' exploration expenditures are a certain percent of the previous period's net profits. Rather than investigating each of the models in detail, the following discussion focuses on motivation and differences in approach.

Perhaps the first widely publicized econometric study of oil and gas exploration was by Fisher (1964). Fisher, who noted that the supply of exploratory drilling differs greatly from the supply of new discoveries, used a three-equation model to attempt to explain (1) the annual number of wildcat wells drilled, (2) success ratio (proportion of wildcat wells drilled which resulted

in a discovery), and (3) the average size of the discovery. Fisher distinguished between exploration at the intensive and extensive margins. Exploration at the extensive margin yields discoveries that are characterized by a relatively low frequency and a large size, whereas exploration at the intensive margin yields relatively small discoveries that occur with greater frequency. Consequently, Fisher specified both the success ratio and the discovery size as a function of economic variables and also asserted that short-term reaction to increases in price results in a shift of exploration to the intensive margin.

In the three-equation model, the number of exploratory wells drilled is specified as a function of the price of oil, geophysical crew time, and lagged values of the average size of oil and gas discoveries, depth, the success ratio, and regional dummy variables. The equation for the success ratio includes the price of oil, geophysical crew time, and lagged values of the success ratio, depth, and average size of oil and gas discoveries. For the final equation, the average-size oil discovery is specified as a function of the price of oil and the previous period's values of the average size of oil and gas discoveries and the success ratio. In order to increase the number of degrees of freedom, short-time series data for large, geologically heterogeneous areas were pooled. Intercept dummy variables were included to account for differences in individual areas. Although the model specification was not derived from a firm's decision process, the explanatory power of the estimated equations was adequate. For predictive purposes, however, the price responsiveness of the new discoveries is unlikely to be the same for all the areas from which the data were taken. Furthermore, variations in the historical price data used by Fisher were the result of differences in the quality of the oil rather than incremental production costs. Erickson and Spann (1971), working to model the supply of natural gas, elaborated on Fisher's original formulation by including an equation for the average annual size of gas discoveries. The functional form of the gas supply model is similar to the functional forms used by Fisher in modeling oil supply.

MacAvoy and Pindyck (1973, 1975) modeled and integrated the components of natural gas demand and supply. Like Fisher, they modeled the supply of new discoveries by predicting the number of wells drilled, success ratios for oil and gas exploration, and the expected size of oil and gas discoveries for individual Petroleum Administration Districts. The supply of reserves may be calculated from the success ratio, wells drilled, and expected size. Both the success ratio and the expected discovery size represent (decline) extrapolations from a calculated reference size and success

ratio. These variables are also sensitive to field prices and signify whether firms are operating at the intensive or extensive margins. The well-drilling equation is based upon the expected returns and the variance of returns which are, in turn, calculated from the expected size of discovery and the success ratio. In particular, the annual number of wells drilled is a function of the petroleum district dummy variables, expected returns, the variance of returns, and a drilling cost index. Although recent revisions (Pindyck, 1978) of the model are logically similar to previous versions, the success ratio and expected size of oil discoveries were made independent of price and dependent upon regional physical characteristics. These revisions led to reductions in previous estimates of the price responsiveness of new oil discoveries.

Two other approaches to modeling new discoveries were made by Khazzoom (1971) and Epple (1975). Khazzoom (1971) estimated the volume of gas discovered from the two-period ceiling averages for the price of natural gas, price of crude oil, price of natural gas liquids, and previous period's volume of gas discovered. The model does not explicitly include a variable associated with the exhaustion of undiscovered deposits. Using a unique approach, Epple (1975) considered exploration as a production process that used wells and oil-bearing land as inputs. Exhaustion of the oil-bearing resource is explicitly considered to be represented by the productivity and input cost of oil-bearing land. The model attempts to predict the value of new discoveries of crude oil and natural gas by using a joint production function. Specifically, Epple assumes a form for the production function,¹ then derives the firm's optimizing conditions under the assumption that it maximizes the net present value of exploration effort. Parameters of the oil-bearing land supply function are estimated from aggregated U.S. oil and gas exploration statistics. Although the analytical approach that Epple takes is novel and some aspects of resource exhaustion are considered, the use of aggregated data and the absence of detail relating to the spatial distribution of deposits restrict the applicability of the model.

Along similar lines, Uhler (1976) developed a stochastic production function for the discovery of new petroleum reserves. The marginal exploration cost function, derived from the production function, includes the following variables: an index of field knowledge, an index of physical exhaustion, and wells drilled. One distinguishing feature of the analysis is that it considers a disaggregated area. Because the study concentrates on the marginal cost of new discov-

eries, it does not attempt to explain the behavioral determinants of wildcat drilling.

Cox and Wright (1976) abstracted data from the exploration process by considering all drilling expenditures as investment in reserves, which are treated as inputs in the production of crude oil. The objective of their study was to link investment in reserves to government policies such as import quotas, prorationing, and special income tax provisions. Although the empirical results are impressive, the authors do not consider exploration investment apart from reservoir development investment. Furthermore, they do not consider the influence of uncertainty or resource exhaustion in their investment decision models.

In summary, economists appear to have taken two approaches to modeling firm exploration behavior. For the first approach, the number of prospects, labor, and capital services serve as inputs to the petroleum reserve production process called exploration. For the second, reserves are considered an input to the crude oil production process. Exploration is then regarded as similar to firm investment in physical capital equipment. To some extent, both approaches are correct, depending upon the nature of the firm under consideration. Firms engaged primarily in exploration regard discovered reserves as the end product, whereas vertically integrated firms might view reserves as inputs. Exploration can be regarded as an investment or input in the production of an inventory of reserves that results in crude oil production. As an intermediate product, reserves will undergo additional modification in the production process in order to generate final products.

OPERATOR BEHAVIOR AT THE FIELD LEVEL

With a few exceptions, the studies just reviewed have been concerned with modeling exploration behavior when data are highly aggregated. Frequently, data limitations necessitate such spatial and temporal aggregation. However, the testable behavioral hypotheses generated from these models are, in general, quite limited. In order to examine operator behavior, in more detail, data specific to the Denver basin are used in this study. In this area deposits were usually found in a single formation; the only exploration play that occurred during the period examined was stratigraphic in nature. Because the initial discoveries were made earlier than the beginning of the historical time series of data used in this study, the behavior examined relates to exploration at the intensive margin. With the exception of the Union Pacific Railroad acreage (Drew and others, 1980), mineral rights were regularly bought and sold. The following discussion concerns the explo-

¹The supply of oil-bearing land is specified as $L = R^B e^{-w}$ where R is unit rent, w is past exploratory effort; with $r > 0$ and $B > 0$ and defined as estimated parameters and e is the Napierian constant.

ration decisionmaking of firms, specifically, how they formulate and respond to expectations of undiscovered deposits remaining in the basin.

FIRM BEHAVIOR

Investment in production of petroleum reserves may take several forms. Additional reserves can be developed by extension drilling (leading to revised estimates of reserves), investment in improved recovery methods (enhanced recovery) from existing reservoirs, and wildcat drilling for the discovery of new deposits. Within the individual firm there are tradeoffs in costs that must be considered when using any one of these three sources of additional petroleum (White and others, 1975). For example, development drilling can be done with borrowed funds, whereas only equity capital can be used to pay for exploratory drilling.

The theory of firm behavior under production uncertainty has only recently been treated extensively in the economic literature (see Fama, 1972; Leland, 1974). If it is assumed that firm managers attempt to maximize the net present value of the firm (perhaps in terms of outstanding capital stock), a capital market valuation model might be posited. For purposes of explaining overall firm exploration decisions, such a model may be particularly applicable because generally only equity capital is used for exploration expenditures. Among the alternatives available to the firm for obtaining additional reserves—that is, wildcat drilling, extension drilling, and enhanced recovery—wildcat drilling clearly is the most uncertain. Firms commonly restrict their exploration activities to a small set of areas where they have had previous experience or where the firm plans a prolonged exploration program. For a given area, early exploration investment can be viewed as investment in the current set of prospects that provides additional information about other prospects in the area. The decision regarding when and in what areas to initiate an exploration program and at what times to allocate expenditures between exploration, extensions, and enhanced recovery are made at the firm level. However, the decision on when and where to drill strategic prospects in an area is frequently the responsibility of the field managers (Kaufman, 1963). This description of the optimizing process of a firm is non-technical. For a technical discussion of the optimal exploration, production, and capital investment policy when the firm operates in geologically diverse regions, see Attanasi (1978).

At the field level, the operator is faced with the problem of allocating a fixed amount of funds to a number of prospects over a given planning period. For larger firms operating in many areas, a marginal amount of exploration funds might be reallocated within the typi-

cal planning period. The optimal search policy for a given set of prospects may be expressed as a function of the expected profit, marginal opportunity costs, and expectations associated with the size (and spatial) distribution of remaining deposits (Attanasi, 1978). Field expectations are formed when operators make exploration decisions based upon unknown or uncertain parameters. For the shortrun model considered here, unobservable operator field expectations appear to be the most difficult components of the determinants of exploration to measure.

FORMATION OF EXPECTATIONS

Expectations are formed when economic agents (firm managers or consumers) are required to make decisions based upon the unknown value of a particular variable. The expectations or predictions that are formed are then used in the decision process. In tests of the theory of the formation of expectations in economic models, the modeler may attempt to gather data directly from economic decisionmakers or construct a prediction scheme based upon observable characteristics of the specific decision problem such as market prices or output. Traditionally, these prediction schemes, although typically rather simple, rely almost entirely upon past values of the variables to be forecasted. A serious limitation of nearly all the existing expectations models is that their estimates are formed without consideration of the decisionmaker's criterion function. Embodied in the decisionmaker's criterion function are attitudes toward risk and consequences of under- and over-estimating the value of the uncertain variable. Originally, expectations models were developed in connection with agricultural price supply responses (Ezekiel, 1938; Nerlove, 1958). Farmers must commit land and labor to the production of specific crops before the market price of the crop is known. The farmer may base his predictions of the current period's price on the previous period's realized price. In particular, suppose p_t^* is the expectations variable and \hat{p}_t is the observed value at time t . The static expectations formulation is defined as

$$p_t^* = \hat{p}_{t-1}. \quad (1)$$

If the current period's expectation is a weighted combination of the realization and expectation of the previous period, that is,

$$p_t^* = p_{t-1}^* + \alpha (\hat{p}_{t-1} - p_{t-1}^*) \quad 0 < \alpha < 1, \quad (2)$$

then expectations are said to be formed adaptively. In the equation above, α is the coefficient of adaptation and controls the degree to which expectations will con-

form to the previous period's realization. Expectations are said to be regressive if

$$p_t^* = a_1 \hat{p}_{t-1} + a_2 \hat{p}_{t-2} + a_3 \hat{p}_{t-3} + \dots, \quad (3)$$

with $a_1 > a_2 > a_3$ the weights are positive and approach zeros with the passage of time. Alternatively, expectations are extrapolative, if

$$p_t^* = a_1 (\hat{p}_{t-1} - \hat{p}_{t-2}) + a_2 (\hat{p}_{t-2} - \hat{p}_{t-3}) + \dots, \quad (4a)$$

and

$$p_t^* = a_1 \hat{p}_{t-1} + (a_2 - a_1) \hat{p}_{t-2} + (a_3 - a_1) \hat{p}_{t-3} + \dots, \quad (4b)$$

where the weights a_i are also positive and decline. In terms of the final coefficients of the lagged terms of p_t , the coefficients (equation 4b) are negative and approach zero. Turnovsky (1969) has shown how a Bayesian updating scheme for uncertain values of the distribution parameters of the stochastic decision variables results in a general expectations formation scheme yielding as special cases static, extrapolative, and adaptive expectations.

As suggested by these formulations and others frequently found in the literature, the estimated form of the unobservable expectations variable is based on the autoregressive process² of the historical realizations of the decision variable. In fact, a relatively unrestrictive definition of rational expectations is that the generated predictions and realizations follow the same autoregressive scheme. A reason for this definition is that autoregressive forecasting schemes provide optimal predictors for a wide range of stochastic processes. However, these predictors are still constructed without recourse to the decisionmaker's criterion function. Moreover, the criterion functions for the prediction and decision problems will, in general, differ because the consequences of errors in the estimates differ. Distributed lag models are frequently applied when the unobservable expectations variable is approximated by the autoregressive formulation of the historical realizations of the variable of interest. If it is assumed that the decisionmakers have knowledge of the underlying stochastic structure that generates the time series of the variable to be forecasted, then the distributed lag proxies are equivalent to calculating the conditional expected value of the uncertain variable (Pesando, 1976).

DISTRIBUTED LAG MODELS

Unobservable expectations variables are frequently modeled with a distributed lag function of the histori-

cal realizations of the variable of interest. Simply, distributed lag models provide the basis for analyzing the temporal response of the dependent variable (y_t) to a change in a specific independent variable (x_t). In particular for any given T , it is assumed that

$$y_t = c + w_0 x_t + w_1 x_{t-1} + \dots + w_T x_{t-T} + e_t, \quad (5)$$

where $\frac{\partial y_t}{\partial x_{t-i}} = w_i \quad i = 0, 1, \dots, T$

and e_t is a stochastic error term.

Two fundamental problems arise if one were simply to apply ordinary least squares to equation (5) to estimate all w 's. First, if T is large, there may be too many parameters, leaving few degrees of freedom to make statistical inferences about parameter values with any degree of confidence. Second, there is likely to be a high degree of collinearity among the lagged values of the independent variables. Consequently, in order to conserve available degrees of freedom and improve efficiency of the statistical estimates, a lag generating function is chosen to correspond to a given time profile that characterizes the response of the dependent variable to changes in a particular independent variable. By assuming a specific analytical form for the generating function, the number of parameters required to be estimated can be made quite small. For example, suppose the decisionmaker's expectation x_t^* is adjusted adaptively as an observation of a realization \hat{x}_{t-1} is made, that is,

$$x_t^* - x_{t-1}^* = \alpha (\hat{x}_{t-1} - x_{t-1}^*) \quad (6)$$

where $0 < \alpha < 1$. If λ is defined by $\lambda = 1 - \alpha$, then the expectations variable may be calculated by the following equations:

$$\begin{aligned} x_t^* &= (1-\lambda) \hat{x}_{t-1} + \lambda x_{t-1}^* \\ x_t^* &= (1-\lambda) [\hat{x}_{t-1} + \lambda \hat{x}_{t-2} + \lambda^2 \hat{x}_{t-3} + \dots]. \end{aligned} \quad (7)$$

Hence, given the derived structure, a single parameter λ can be used to generate values of the unobservable expectations variable from the historical data associated with x_{t-1} . Suppose it is assumed that

$$y_t = x_t^* \quad (8)$$

Then recursively substituting (7) into (8) and the lagged form of (8) into the resulting equation, the estimation form of equation (8) is

$$y_t = (1-\lambda) \hat{x}_{t-1} + \lambda y_{t-1} \quad (9)$$

² The term autoregressive process used here refers to a model that can be expressed as a finite linear aggregate of previous values of the process. The nature of the stochastic process might correspond to an autoregressive model, moving average model, or combination of the two models. For more detailed definitions, see Box and Jenkins (1970, p. 7-11).

or using the lag operator L the equation is

$$y_t = \frac{(1-\lambda)\hat{x}_{t-1}}{(1-\lambda L)}.$$

Consequently, by deriving the specific form of the lag generating function, the estimation of the distributed lag structure of the model can be reduced to a small number of parameters.

Jorgenson (1966) has shown that any arbitrary lag function can be approximated by the rational form³

$$y_t = \frac{A(L)}{B(L)} x_t \quad (10)$$

$$\text{where } A(L) = \sum_{i=0}^m a_i L^i \quad B(L) = \sum_{j=0}^n b_j L^j.$$

A rational lag function means that the function can be expressed as the ratio of two polynomial (lag) functions. That is, any finite or infinite response function can be approximated by the rational form (10). For example, a form of the geometric function results when $A(L) = \alpha I$ and $B(L) = 1 - \lambda L$, that is,

$$y_t = \frac{(\alpha I)}{(1-\lambda L)} x_t. \quad (11)$$

Alternatively, by specifying the denominator in (10) to be equal to a constant (unity), a finite distributed lag function results. In general $B(L)$ of higher degree than 2 or 3 might be difficult to identify uniquely. The statistical methods used to estimate such models are complex because the estimation routines are highly nonlinear. In addition, the distribution of the descriptive statistics are known only for large samples (Dhrymes, 1971).

In terms of temporal behavior, the implications of particular distributed lag schemes are not always obvious from the estimated structural equation form. For model comparison and selection, certain standard properties of the implied function, which are more intuitively interpretable, are frequently considered. Two such properties are the normalized time profile of the distributed lag response function and the average lag length. The time profile represents the temporal distribution of the effects on the dependent variable of a unit change in the independent variable. The general shape of the time profile indicates how rapidly and in what direction a change in the independent variable affects the dependent variable. Changes in the units of measurement of either the dependent or independent vari-

ables affect the shape of the time profile. Consequently, when the time profiles for alternative distributed lag models are compared, they must be taken to use the same units of measurement. The time profile may exhibit a variety of forms (fig. 1). Some common ones include an exponentially decaying function, an inverted "V" shape or a function with weights oscillating between positive and negative values. In some cases, the general shape of the time profile can be inferred from the functional form of the distributed lag model. However, the time profile weights can always be calculated using the estimated function and tracing the effects on the dependent variable of a unit change in the specific independent variable.

Another property of distributed lag functions that may serve as a basis for comparison is the average lag for an n period lag function, which is calculated as

$$\Theta = \frac{\sum_{t=0}^n tw_t}{\sum_{t=0}^n w_t}, \quad (12)$$

where t is the period subscript and w is the associated weight. The average lag Θ turns out to be a simple weighted average of the time periods where the weighting is proportional to the time profile weights. The average lag reflects how the various values of the time profile weights are temporally distributed. Values of the average lag must be interpreted in relative terms. That is, larger values of Θ indicate that much of the distributed lag effect is felt at larger values of t , and smaller values of Θ indicate that the weights associated with earlier lag periods are relatively larger than for the later ones.

The time profile and average lag represent properties of the distributed lag model that are easily interpreted. Frequently, an individual researcher may have subjective prior knowledge of the form for the time profile and attempt to use this prior knowledge to evaluate the empirical model. For example, suppose a distributed lag model is estimated representing the effects of increases in advertising on product sales. One might reasonably expect the time profile to indicate the effect of a one-time increase in advertising to be exponentially decaying, but it would be unreasonable to expect the function to indicate explosive or monotonically increasing effects. The time profile provides a means whereby the estimated distributed lag model can be evaluated for consistency with expected economic behavior.

OPERATOR DRILLING BEHAVIOR

Assuming the type of decentralized decisionmaking that was previously described, the field manager's re-

³The general form of the rational distributed lag function may be derived from an extension of the adaptive expectations hypothesis (see Dhrymes, 1971).

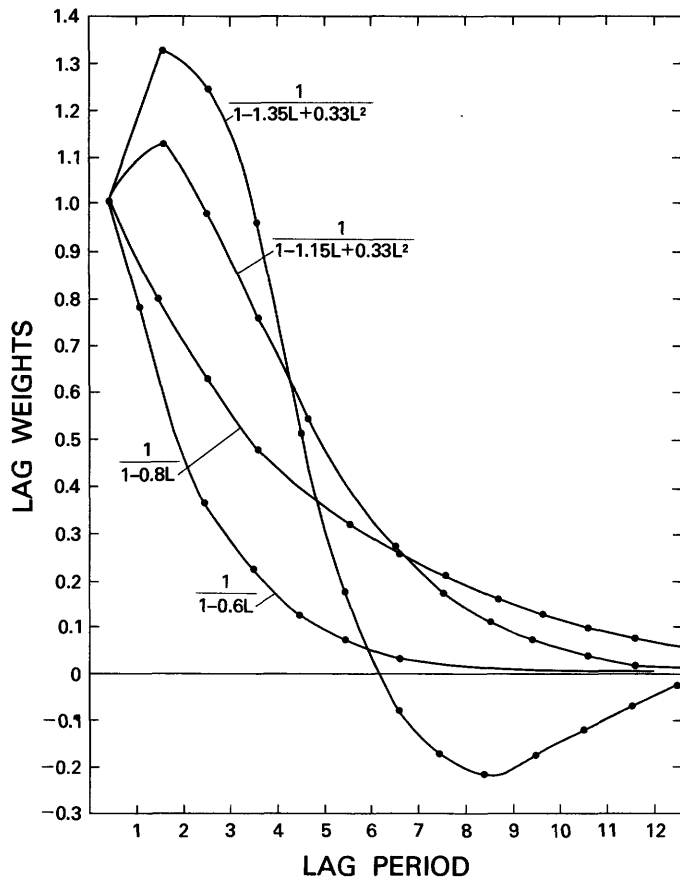


FIGURE 1.—Shapes of normalized time profiles for weights associated with values of the lag variable. Analytical specifications of the distributed lag functions are adjacent to each of the time profiles with L defined as the lag operator.

source allocation problem might then be specified as a sequential decision or adaptive control problem. Without going into an involved mathematical derivation, the optimal shortrun search strategy might be specified as a function of profit expectations, perceived field risks, and expected opportunity costs of foregone alternatives. Field managers are assumed to use available information to predict the spatial and size distribution of undiscovered deposits. The optimal search effort might be specified as a function of expected profit $\hat{\pi}$ and expectations associated with deposits left to be found χ^* . Depending on the nature of the area, a variable k might also be included to index the current state of field knowledge relating to the original distribution of deposits and extent of physical exhaustion. The optimal search effort may be specified by the following function:

$$y_t = f(\hat{\pi}_t, \chi_t, k_t). \quad (13)$$

To estimate a specific functional form of equation (13), the variables $\hat{\pi}_t$ and k_t must be defined. It seems reasonable to assume that there is a lag between the

firm's exploration allocation decisions across regions and the reporting of profits from which π_t is assumed to be estimated. In the empirical models to be presented, the expected profits variable was computed as the lagged value of returns per wildcat well. Returns per exploratory well were calculated by multiplying the quantity of the dollar value of oil found per successful wildcat less the cost of the well by the success ratio. This variable was assumed to reflect the general level of a basin's exploration profitability. The variable k_t is assumed to reflect the relative degree of knowledge operators possess about a particular set of targets and the extent of physical exhaustion. This relative index of experience and exhaustion k was calculated on the basis of the weighted average of the cumulative number of wells drilled in a specific target area over the previous three periods.⁴ The index as defined would have greater significance and intuitive appeal if the area under consideration exhibited several plays or target formations. Because a large block of acreage in the Denver basin was restricted from exploration, the index was calculated on the basis of areal extent withheld rather than on the basis of target formation (see footnote 4). The index specified in this fashion captures the effects of sampling deposits without replacement and also will increase as economically exploitable deposits are exhausted. The wildcat drilling data and reserves data were obtained from the Well History Control File of Petroleum Information, Inc. Drilling costs were calculated from various annual issues of the Joint Association Survey of the U.S. Oil and Gas Producing Industry (JAS) (American Petroleum Institute, 1953, 1955–56, 1959–73). The study area was partitioned into 88 smaller units (of 625 mi² or 1,619 km²) for purposes of assigning target depths for wildcat wells drilled within each cell. Well costs were calculated from the inferred target depth, and costs per foot were taken from the JAS annual summaries for Colorado. Although some natural gas deposits were found, operators were searching predominantly for new oil deposits. The oil found was of a relatively uniform quality. During the period from 1949 to 1972, prices were relatively stable and ranged from \$2.63 per barrel to \$3.46 per barrel. However, during 1973 prices for new oil increased to over \$8.90 per barrel. Initially, the general estimated form of equation (13) was

$$y_t = \beta_0 + \beta_1 \hat{\pi}_{t-2} + f(D_t) + \gamma k_t + \epsilon_t, \quad (14)$$

⁴If P_i is the proportion of wells drilled in the i^{th} formation or sub basin at time t and C_i is the cumulative number of wells drilled in the i^{th} formation or subarea, then the experience-exhaustion index is defined by $k_t = \sum_{i=1}^m C_i \cdot (1/3) \left(\sum_{j=1}^3 P_i^{t-j} \right)$. The Denver basin had two subareas and a single producing formation. The areas were partitioned into acreage excluded from exploration and held by the Union Pacific Railroad; other acreage corresponded to the rest of the basin which was open for exploration.

where $f(D_t)$ represents the general lag structure and $\chi_t^* = f(D_t)$. The variable y_t represents search effort expended as measured by either wildcat wells drilled or estimated drilling expenditures, $\hat{\pi}_{t-2}$ is the expected value of returns per wildcat well lagged two periods, D_t is the value of new discoveries equal to or greater than a given minimum size class of deposit for a specific time period, and k_t is the index of field experience described earlier. The lag function of the dollar value of the oil found in certain size classes of deposits per time period is assumed to reflect changes in the operator's expectations regarding the distribution of deposits remaining to be found in the basin. The variable D_t was defined with respect to a minimum size class of deposits for two reasons: Larger deposits generally have lower production costs; moreover, the occurrence of larger deposits is frequently taken to be indicative of the presence of deposits of a similar size.

Estimated empirical models are presented in the following two sections. In the first section, alternative forms of the lag function were tested; the final model took the form of a finite lag function with lag weights constrained to lie in a second-order polynomial. Inferences about operator behavior are made from the estimated models. On the basis of these results, inferences about operator behavior observed in the Denver basin are compared with inferences about behavior observed in the Powder River basin (Attanasi and Drew, 1977). In the next section, the estimated quarterly model is compared to the model estimated from semiannual observations for consistency in interpretation. The discovery process model similar to the model presented in Arps and Roberts (1958) is then used to calculate the value of deposits remaining to be found. The performance of this type of model and its apparent efficiency in the use of information contained in the drilling data suggest that the model predictions are equivalent to the expected value of the uncertain variable. A more restrictive definition of a rational expectation is that it efficiently uses all available information about the uncertain variable. It would therefore be useful to compare the statistical properties of the ad hoc distributed lag model with the performance of an empirical drilling model that uses the (rational) predictions of the value of deposits remaining to be found calculated with the discovery process model.

EMPIRICAL RESULTS

Several forms of the distributed lag function $f(D_t)$ were fit to the historical drilling data in order to arrive at a reasonable description of the discovery process. In evaluating alternative empirical models, several criteria are frequently applied. The model structure and estimated coefficients should first be consistent

with economic theory. Other criteria relate to the standard statistics that describe the quality of the fit of the model, that is, coefficient "t" statistics, coefficient of determination and the standard error of the regression. Finally, the predictive performance of the model can be examined by using part of the historical sample to estimate the model and the remainder of the sample to compare model forecasts with actual sample values.

The specific lag forms that were tested included the geometric, second-order rational, and finite polynomial distributed lag functions. The geometric lag structure is derivable if it is assumed that expectations are formed adaptively.⁵ If, in addition to an adaptive expectations hypothesis, it is assumed that reactions to changes in expectations exhibit some inertia—that is, only a partial immediate adjustment (see Johnston, 1972, p. 302–303)⁶—a second-order rational lag model results. Motivation for the finite polynomial lag function was not based on any specific set of assumptions that explain the process by which expectations are formed. However, the empirical results for the finite lag distribution will provide a useful comparison to the results obtained for alternative infinite lag forms, that is, the geometric and second-order rational functions. An infinite lag distribution means that the effect on the dependent variable of a change in the lag variable continues over an infinite number of time periods.

The parameters for the geometric and second-order lag functions were obtained using a nonlinear iterative regression technique that yields maximum likelihood estimates (Pierce, 1971). For the initial test runs, D_t was defined as the aggregate value of discoveries per time period of deposits of at least 500,000 barrels of reserves. Prior to carrying out the estimation procedures, the data were tested for seasonal variations, which were removed from the quarterly data using the Census X-11Q procedures (Shiskin and others, 1967). Table 1 presents the parameters estimates for models based on semiannual and quarterly observations. Similar results were obtained when $\hat{\pi}_{t-1}$ was used in place of $\hat{\pi}_{t-2}$ and when alternative assumptions were imposed on the error term. Comparisons of the coefficients of the lag terms associated with the models estimated from

⁵For this derivation see the preceding discussion and Attanasi and Drew (1977, p. 57). The estimated form of the geometric lag function was

$$y_t = \beta_0 + \frac{\beta_1 D_t}{(1 - \lambda L)} + \beta_2 \hat{\pi}_{t-2} + \gamma k_t + \epsilon_t,$$

where y_t , D_t , $\hat{\pi}_{t-2}$, k_t were defined earlier and β_0 , β_1 , λ , β_2 , and γ are estimated coefficients. L is the lag operator. The stochastic component of the model is embodied in ϵ_t and was permitted to have autoregressive and moving average components.

⁶The second-order rational lag function that was estimated took the following form:

$$y_t = \beta_0 + \frac{\beta_1}{(1 - \lambda_1 L + \lambda_2 L^2)} D_t + \beta_2 \hat{\pi}_{t-2} + \gamma k_t + \epsilon_t,$$

where the variables y_t , D_t , $\hat{\pi}_{t-2}$, k_t were defined earlier and the estimated coefficients are β_0 , β_1 , β_2 , λ_1 , λ_2 , and γ . L is the lag operator. The error term ϵ_t was assumed to have autoregressive and moving average components.

TABLE 1.—*Estimated coefficients of infinite lag models*[Numbers in parentheses are "t" statistics, and \bar{R}^2 is the coefficient of determination]

Geometric:							
Drilling:							
Semiannual	----	-488.0	0.00102	0.852	-0.0173	0.066	0.735
		(4.1)	(8.2)	(52.5)	(.8)	(5.8)	
Quarterly	-----	-17.5	.00039	.899	-.0052	.0098	.806
		(1.4)	(11.7)	(100.2)	(1.0)	(6.7)	
Drilling expenditures:							
Semiannual	----	-2768.0	.00450	.881	-.0690	.320	.766
		(4.6)	(7.9)	(68.8)	(.7)	(5.7)	
Quarterly	-----	-40.4	.00161	.923	-.0241	.0258	.830
		(.7)	(12.2)	(130.6)	(1.0)	(3.6)	
Second order:							
Rational Lag							
Drilling:							
Semiannual	----	-394.4	0.00064	1.284	-0.389	-0.0266	0.0589
		(3.5)	(3.9)	(7.9)	(2.6)	(1.3)	(5.4)
Quarterly	-----	-15.3	.00019	1.491	-.543	-.00324	.0101
		(1.3)	(4.0)	(10.2)	(3.8)	(.7)	(7.6)
Drilling expenditures:							
Semiannual	----	-2370.0	.0028	1.270	-.354	-.100	.290
		(3.9)	(3.1)	(5.8)	(1.8)	(.1)	(5.0)
Quarterly	-----	-28.4	.00096	1.363	-.411	-.0192	.0262
		(.5)	(3.9)	(7.0)	(2.2)	(.8)	(3.6)

the semiannual and quarterly data show surprising similarity. Parameter coefficient estimates of the expected profits and the field exhaustion variables which are negative and positive, respectively, are inconsistent with prior theoretical expectations. The negative expected profit coefficient is, in part, due to multicollinearity between $\hat{\pi}_t$ and D_t . Because the empirical estimates were inconsistent with the theoretical restrictions, both infinite lag forms were rejected.

There seems to be no a priori reason for restricting the distributed lag function to an infinite lag distribution. A finite lag model was estimated in which the lag weights or coefficients were constrained to lie within a polynomial function (Almon, 1965). As discussed earlier, behavioral interpretations associated with the derivation of such lag models are limited. The form of the model used to study wildcat drilling expenditures is

$$y_t = \beta_0 + \beta_1 \hat{\pi}_{t-2} + \sum_{i=0}^m \omega_{t-i} D_{t-i} + \gamma k_t + \epsilon_t. \quad (15)$$

Coefficients ω_{t-i} are the weights associated with $(t-i)$ th period and reflect the temporal effects of a change in D_{t-i} on operators' exploration effort. Assuming that a finite lag model is appropriate, it is important to determine the appropriate lag length for D_{t-i} and the degree of the polynomial containing the lag weights.

Owing to the presence of serial correlation in the data, the application of ordinary least-squares regression procedures in the linear equations would have resulted in inefficient parameter estimates. The Cochrane-Orcutt procedure, which is described in detail in Kmenta (1971, p. 287-289), was used to generate efficient estimates. This procedure amounts to the iterative application of ordinary least squares on data that have been transformed using an estimate of the serial correlation coefficient ρ . Ordinary least

squares is initially applied to the raw data to obtain an estimate of ρ from the residuals. The data are transformed, and least squares is applied to the transformed data. A new estimate of ρ is obtained from the residuals of the second regression. If x_t is the t th observation of the variable, then the transformed variable \bar{x}_t is defined by $\bar{x}_t = x_t - \rho x_{t-1}$. This iterative procedure continues until there is no significant change in ρ .

The consequence of understating or overstating the true lag structure is the introduction of a specification error (Schmidt and Waud, 1973). If the finite lag model weights are interpreted as representing the temporal response of operators' search effort to new discoveries, then certain restrictions should be placed on the lag weights to be consistent with economic theory. For example, the weights would be difficult to interpret in this fashion if they oscillated between positive and negative signs, or if they monotonically increased as the lag period became more distant. Given that the pattern of the coefficients conforms to certain prior restrictions, Schmidt and Waud (1973) suggest that appropriate lag length be determined after estimating the model for a number of different lengths and that the final choice be made on the basis of maximizing the adjusted coefficient of determination or minimizing the standard error of the regression. In this study, the lag structure was determined by searching over various combinations of lag lengths, and the choice of the final form was based on minimizing the standard error of the regression equation. In this search, D_{t-i} represented the value of new discoveries found in deposits of at least 500,000 barrels. Estimates of coefficients for the finite distributed lag models are presented in tables 2 and 3. For the model based on semiannual data, the appropriate choice of lag length was five periods⁷, with the coefficient for the fifth period restricted to zero. The appropriate lag length for the model based on quarterly observations was found to be 11 periods with the coefficient for the final period restricted to 0. When the initial period is also taken into consideration, the empirically determined lag lengths for the two models (semiannual and quarterly) are consistent. The time profile indicated by the lag weights have the shape of an inverted "V".

Using the semiannual data, the finite lag models were reestimated with D_t redefined to represent the total value of new discoveries of deposits greater than 2.5 million barrels and 5 million barrels. A comparison of the estimated models and associated statistics across

⁷ One problem that arises in determining the appropriate lag length and that has not been addressed in the literature is whether the models of varying lag length should be estimated using the maximum sample size or using identical sample sizes. For the model specifications that were tested, the five-period model estimated from the maximum sample size had the minimum standard error of the regression, and the seven-period model had the minimum error when alternative models were estimated using identical samples.

TABLE 2.—Coefficients of polynomial distributed lag models, semiannual observations

[Specifications A and B are based on 46 observations from exploration activity in the Denver basin. Specification C is based on 29 observations from exploration activity in the Powder River basin. \bar{R}^2 , adjusted coefficient of determination; D.W., Durbin-Watson statistics; S.E., standard error of regression. Numbers in parentheses are absolute values of "t" statistics]

Wells														
A	213.9 (2.2)	0.0340 (.6)	-0.271 (1.1)	0.279 (1.6)	0.579 (3.5)	0.634 (4.1)	0.441 (4.2)	-0.005 (.4)	-----	-----	-----	0.756	2.3	60.5
B	184.6 (1.5)	.0330 (.6)	-.227 (.8)	.309 (1.6)	.600 (3.5)	.645 (4.1)	.445 (4.2)	.005 (.2)	0.005 (.2)	-0.001 (.2)	-0.012 (.2)	.749	2.3	61.4
C	47.5 (5.4)	.0054 (2.4)	-.061 (1.5)	.192 (6.8)	.302 (11.3)	.269 (11.2)	.094 (2.7)	-.053 (5.0)	-.051 (4.9)	-.006 (1.4)	.117 (5.1)	.940	2.1	13.6
Expenditures														
A	2098.0 (3.2)	0.5180 (1.8)	-1.420 (1.0)	-0.072 (.1)	0.744 (.7)	1.028 (1.1)	0.780 (1.2)	-0.165 (2.1)	-----	-----	-----	0.756	2.6	325.8
B	1577.0 (2.1)	.5090 (1.7)	-.976 (.7)	.272 (.3)	.994 (1.0)	1.189 (1.3)	.858 (1.4)	.092 (.6)	0.080 (.5)	0.036 (1.5)	-0.257 (.9)	.756	2.4	325.9
C	932.4 (7.8)	.0410 (1.6)	-1.740 (3.8)	.253 (.7)	1.271 (3.8)	1.313 (4.3)	.381 (.9)	-.545 (4.2)	-.571 (4.4)	.077 (1.5)	.936 (3.4)	.764	2.2	160.1

TABLE 3.—Coefficients of polynomial distributed lag models, quarterly data

[\bar{R}^2 , coefficient of determination; D.W., Durbin-Watson statistic; S.E., standard error of regression]

	Drilling		Drilling expenditures	
	Coefficients	"t" statistics	Coefficients	"t" statistics
β_0	112.7	2.2	801.0	2.9
β_1	.0096	1.7	.0275	.9
β_2	-.0044	.7	-.0546	1.7
ω_0	.017	.3	.126	.4
ω_1	.110	2.7	.401	1.3
ω_2	.184	3.2	.620	2.1
ω_3	.240	4.0	.781	2.4
ω_4	.276	4.3	.884	2.5
ω_5	.293	4.4	.930	2.6
ω_6	.292	4.4	.919	2.6
ω_7	.271	4.4	.850	2.6
ω_8	.232	4.4	.724	2.5
ω_9	.174	4.4	.540	2.5
ω_{10}	.096	4.4	.299	2.5
\bar{R}^2	.912		.912	
D.W.	1.88		1.98	
S.E.	17.45		94.89	

these deposit size categories indicated little difference in terms of the overall model form or the explanatory power of the model specifications.⁸ It might reasonably be concluded that operator expectations regarding the value of remaining deposits were sustained by the quantity of oil found per time period rather than by the discoveries of specific size deposits. This finding is consistent with results of a similar study carried out for the Powder River basin (Attanasi and Drew, 1977). Both the profit and field experience exhaustion variables are statistically more significant in the expenditures equation than in the drilling equation.

In order to determine if the experience-exhaustion index exhibited systematic intertemporal effects, a lag structure was imposed on the variable k_t , and the equation was reestimated. Results of these specifications are presented in table 2. None of the lag coefficients was significant in both the drilling and drilling expenditures equations, nor was there any improvement in the explanatory power of the equations.

⁸ The adjusted coefficients of determination (\bar{R}^2) for a five-period lag when the minimum size classes of deposits are defined as 2.5 million and 5 million barrels are 0.7434 and 0.7455 for well drilling and 0.7131 and 0.7550 for well drilling expenditures, respectively.

Neither drilling or drilling expenditures are perfect measures of exploration effort. Drilling activity, although easily interpreted in terms of sampling effort, cannot differentiate between the wells drilled under varying depths or qualitative degrees of difficulty. Although drilling expenditures may more completely reflect the economic dimension of search effort, they cannot fully represent the temporal behavior because the costs of land acquisition and preliminary geophysical work are not included. The negative coefficients associated with the first periods of the distributed lag models based on semiannual observations appear to be the result of two factors. First, drilling and drilling expenditures do not reflect the proper magnitude of search effort in the early stages of exploring a prospect. Second, adjustments in onshore exploration activity would be more readily captured if observations were associated with shorter time periods, as shown by the model based on quarterly data.

The estimated responses can also be compared to results obtained for exploration activity in the Powder River area (specifications C, table 2). In the study of that area, the geometric, second-order rational lag and finite lag models were also estimated. The model that most appropriately described the data was again found to be a finite lag distribution. The empirically determined lag length for the Denver basin was five semiannual periods with the fifth-period coefficient constrained to zero. The appropriate lag structure for the Powder River basin was determined to be four periods with the coefficient for the fourth period unconstrained. Major differences between the estimated response functions are related to the differences in the magnitudes of the lag function weights and the significance of the field experience-exhaustion variable. The lag coefficient on the expenditures equations are comparable. However, the values of the lag coefficients for the drilling equation for the Powder River basin area are about half of those for the Denver basin.

This result, in part, reflects the fact that wells were two or three times as expensive to drill in the Powder River basin. Because there were three exploration plays and four different target formations in the Powder River basin, the field experience-exhaustion index varied widely. The estimated coefficients are consistent with this observation and indicate that exploration effort in the Powder River basin was significantly influenced by the current state of experience or exhaustion. Differences in the estimated relations appear to be reasonably consistent with differences in economic and geologic factors that are reflected in the specific basin's exploration history.

In order to compare the implied temporal properties of the alternative lag schemes, the time profiles of the geometric, second-order rational lag and polynomial (finite) distributed lag schemes are presented in figures 2 and 3. Infinite lag functions consistently overesti-

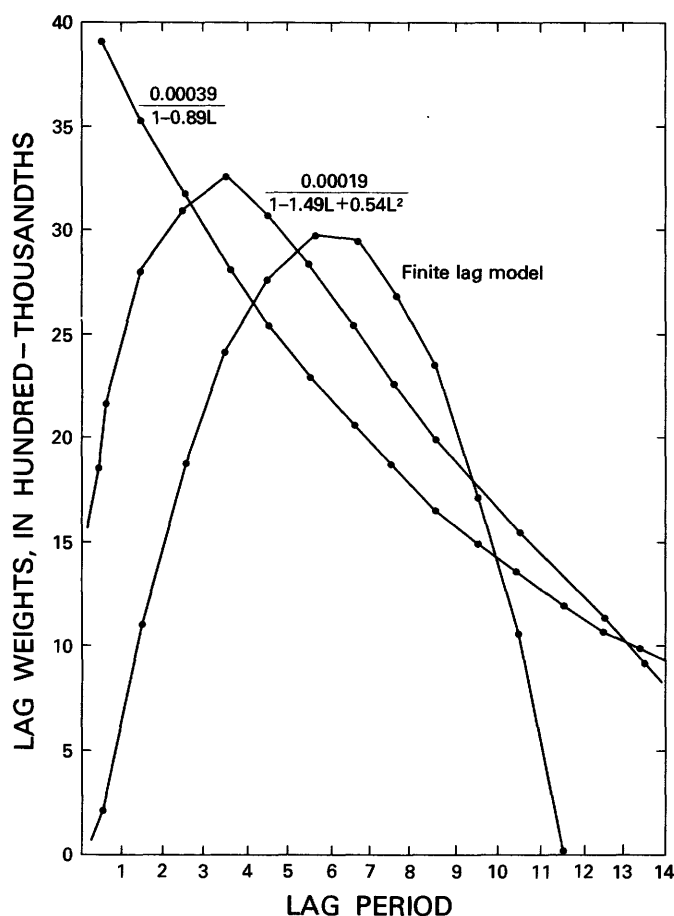


FIGURE 2.—Temporal effects of a one-time unit change in the value of total discoveries (in thousands of dollars) of at least 500,000 barrels on wildcat drilling. Specifications of the lag function are adjacent to each of the time profiles with L defined as the lag operator.

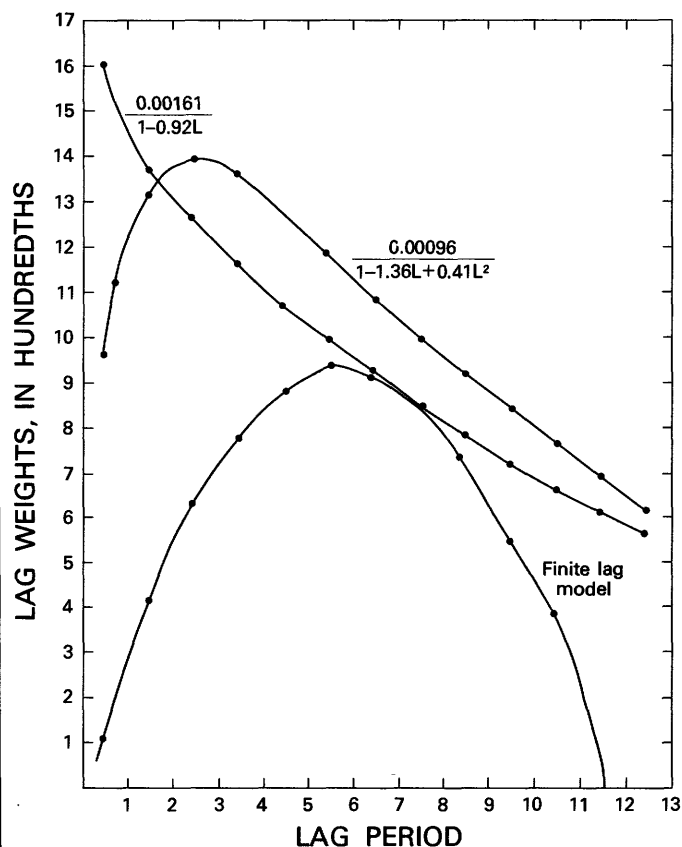


FIGURE 3.—Temporal effects of a one-time unit change in the value of total discoveries (in thousands of dollars) of at least 500,000 barrels on drilling expenditures. Specifications of the lag function are adjacent to each of the time profiles with L defined as the lag operator.

mate operators' responses⁹ because, in the infinite lag models, the negative coefficients on the profit variables (table 1) inflate the role of the lag function in inducing exploration effort.

APPLICATION OF THE DISCOVERY PROCESS MODEL PREDICTIONS

A fundamental premise of the approach to field exploration modeling taken here is that operators are induced to explore as long as there are high expectations associated with the value of remaining deposits. In order to test this premise, a discovery process model was used to provide predictions of the undiscovered recoverable resources as a function of the cumulative number of wells drilled. Two discovery process models that are very similar in their initial assumptions and that have been estimated for the Denver basin are the models described in Arps and Roberts (1958) and Drew, Schuenemeyer, and Root (1980).

⁹ These comments are relevant if the distributed lag model is interpreted to represent operators' temporal response functions to new discoveries.

Both models proceed from the assumption that the probability of finding a deposit in a particular size class is proportional to the number of undiscovered deposits of that size and proportional to the ratio of the area of the typical deposit of the deposit class to the relevant area of the basin under exploration. The model of Arps and Roberts (1958) specifies that the rate of discovery (within a size class of deposits) declines exponentially as a function of the number of exploratory holes drilled. Alternatively, Drew, Schuenemeyer, and Root (1980) use a somewhat different functional form to express the decline rate. They are able to estimate objectively the discovery efficiencies and relevant basin size, whereas Arps and Roberts used prior subjective information to make these estimates. The discovery process model described below is similar to the Arps-Roberts model in functional form, although the data used to estimate the model are more recent and accurate. Because of its computational simplicity, the model of Arps and Roberts was used instead of the model by Drew, Schuenemeyer, and Root (1980). Both models appear to predict the historical discovery sequence quite well (Drew, Schuenemeyer, and Root, 1980).¹⁰

One function of the discovery process model is to provide a framework for predicting future recoverable resources as a function of the cumulative number of wells drilled. An important component of the model is the estimate of the ultimate or total number of deposits within the particular size class of deposits, given as $F_i(u)$ where i denotes the deposit size class and u is an arbitrary large number of wells (perhaps infinite) that exhausts or finds all the deposits in the i th size class. This estimate positions the extrapolation of the discovery decline curve. Given the parameters of the discovery process model, historical data may be used to obtain an estimate of $F_i(u)$. During the discovery history of the basin, estimates of $F_i(u)$ may be made and operator exploration behavior tested in order to determine (1) if exploration is induced by the expectations associated with the remaining deposits and (2) if exploration is sustained by expectations attached to specific size classes of deposits or the cumulative value of remaining deposits.

The Arps-Roberts discovery process model has the following form:

$$F_i(w) = F_i(u)(1 - e^{-\theta_i w}) \quad (16)$$

where $F_i(w)$ is the predicted number of deposits within size class i , $F_i(u)$ is the ultimate number of deposits in the class, w is the cumulative number of wildcat wells,

and θ_i is a constant that includes the basin size, the area of the typical deposit from a given size class of deposits, and the exploration efficiency (Arps and Roberts, 1958, p. 2563). The basin size B was taken to be 5,700,000 acres (approximately 23,068 km²) (Arps and Roberts, 1958). The exploration efficiency was assumed to be 2 for all size classes of deposits, and the area A_i associated with a deposit size class i was taken from Drew, Schuenemeyer, and Root (1980).¹¹ More specifically for the i th deposit size class,

$$\theta_i = \frac{2A_i}{B} \quad (17)$$

Although exploration efficiency does vary across size classes of deposits, tests of the model using the constant 2 for all deposit size classes did not produce any appreciable loss in predictive accuracy. Arps and Roberts (1958) indicate that predictions of the model are equivalent to finding the (mathematical) expected value of the number of discoveries to be made with a given search effort. Consequently, when this discovery process model is used to predict the value of undiscovered deposits, the specific forecasts will be referred to as "rational" predictions or "rational" expectations.

The "rational" predictions of the number of remaining deposits for each size class of deposits was generated in the following manner. An estimate of $F_i(u)$ was made by calculating

$$F_i(u) = F_i(w) / (1 - e^{-\theta_i w}), \quad (18)$$

where $F_i(w)$ is the number of deposits already found with w wells within the i th size class. The number of deposits remaining to be found is the difference between $F_i(u)$ and $F_i(w)$. By assuming a current price of oil and an average size to deposits within each size class, the value of oil remaining to be found in the basin is predicted after summing the values across size classes. The estimation value of remaining deposits is

Deposit size in millions of barrels		Area in acres per square kilometer (values in parentheses are in square kilometers)	
0.0 - 0.002	2.500	(.010)	
.002 - .004	3.900	(.016)	
.004 - .008	6.100	(.025)	
.008 - .016	11.700	(.047)	
.016 - .032	19.400	(.079)	
.032 - .064	34.000	(.138)	
.064 - .128	54.700	(.221)	
.128 - .256	97.300	(.394)	
.256 - .500	216.960	(.878)	
.500 - 1.00	328.320	(1.329)	
1.00 - 2.00	481.920	(1.950)	
2.00 - 4.00	697.600	(2.823)	
4.00 - 8.00	1068.800	(4.325)	
8.00 - 16.00	1324.800	(5.361)	
16.00 - 32.00	2752.000	(11.137)	
32.00 - 64.00	6400.000	(25.900)	

¹⁰ Root and Schuenemeyer (1980) show the Arps and Roberts (1958) model and the limit discovery process model of Drew, Schuenemeyer, and Root (1980) to converge in the limit.

updated each period using both the most recent discovery data and the data from previous periods. This scheme for generating "rational" expectations can naturally be interpreted to suggest that each period operators using available information sequentially update their estimates of the value of undiscovered deposits remaining in the basin. Empirical tests that are carried out are related to how closely levels of exploration effort correspond to the predicted value of remaining deposits and whether exploration is correlated more with predictions for specific size classes of deposits or with the cumulative value of deposits remaining.

The models explaining exploration behavior were specified to include the lagged profit term and the predicted value of undiscovered deposits. Because the discovery process model implicitly takes into consideration the effects of physical exhaustion, the variable k_t was omitted from the model specifications. Estimates of $F_t(u)$ can be quite erratic when there are relatively few wildcat wells drilled in the basin. Consequently, the data set that was used included the period from the third quarter in 1951 through 1973. Again, the Cochrane-Orcutt procedure was applied to handle problems of autocorrelation (Kmenta, 1971). The estimated equations and the statistics describing the quality of the estimates are presented in the following equations:

for drilling,

$$y_w = 14.160 + 0.00934\hat{\pi}_{t-2} + 0.10172\chi_t^* \quad (19)$$

(0.4) (1.4) (3.3)

$$\bar{R}_2 = 0.885 \quad \text{S.E.} = 20.72 \quad \text{Durbin-Watson statistic} = 1.45;$$

for drilling expenditures,

$$y_e = -49.010 + 0.02559\hat{\pi}_{t-2} + 0.58802\chi_t^* \quad (20)$$

(0.4) (1.7) (3.6) (3.6)

$$\bar{R}_2 = 0.898 \quad \text{S.E.} = 104.36 \quad \text{Durbin-Watson statistic} = 1.72.$$

In these equations R^2 is the adjusted coefficient of determination, S. E. is the standard error of the regression, and the numbers in parentheses below the coefficients are the associated "t" statistics. The variable χ_t^* represents the total value (across all classes) of deposits remaining to be found in the basin. Regressions using particular size classes of deposits for the variable associated with operators' expectations were also carried out. Although they produced slight differences in explanatory power, the improvements were not significant. From these results it can be concluded that, with reference to exploration, operators appear to react

to the total perceived economic value of undiscovered deposits rather than to the presence of a specific size of deposits.

Predictions generated by the discovery process model could also be used as arguments in a distributed lag model. This use might be rationalized by assuming that the current estimates of the value of undiscovered deposits to which operators react are weighted averages of previous estimates. Several model specifications of this type have been estimated. However, the results indicated no substantive improvement in the statistics describing the quality of the models. In some cases, the signs of the lag coefficients were difficult to rationalize. Consequently, the model estimates are not presented or discussed in the subsequent analysis.

The qualitative statistics that describe the fits of the ad hoc distributed lag model and the model based on the predictions of the discovery process model are comparable. It would be of particular interest to determine if the predictive performances of the models are similar. For a situation of equal reliability, the use of the simpler model is to be preferred since it has a somewhat less ad hoc foundation and appears to use available observations for parameterization efficiently. The motivation for comparing the performance of the models beyond their respective qualitative statistics is to identify systematic relative weaknesses. Procedures for the comparative evaluation of econometric models have only recently been discussed in detail in the literature (Dhrymes and others, 1972). During the process of constructing the present models, several evaluation criteria have already been applied. These criteria concern how well the model structure conforms to accepted economic theory and if the estimates are theoretically consistent. Other techniques compare the pattern of the model predictions with the historical data. One method is to examine whether the model predicts the turning points that actually occur in the dependent variable. In this regard, the performance of both models was relatively poor. Generally, the predicted turning points were lagged one period from the historical turning points. For the well-drilling equations, both models predicted only 2 of the 27 turning points correctly; for the expenditures equations, the "rational" model predicted 5 of 26, and the distributed lag model predicted 3 of 26 correctly. One explanation for the relatively large number of turning points in the data is that the data appear to contain a residual seasonal variation that was not removed.

Another technique frequently used in model evaluation is to examine the pattern in the residuals. Figures 4 and 5 plot the residuals of the two models over a common sample period. The plotted residuals do not

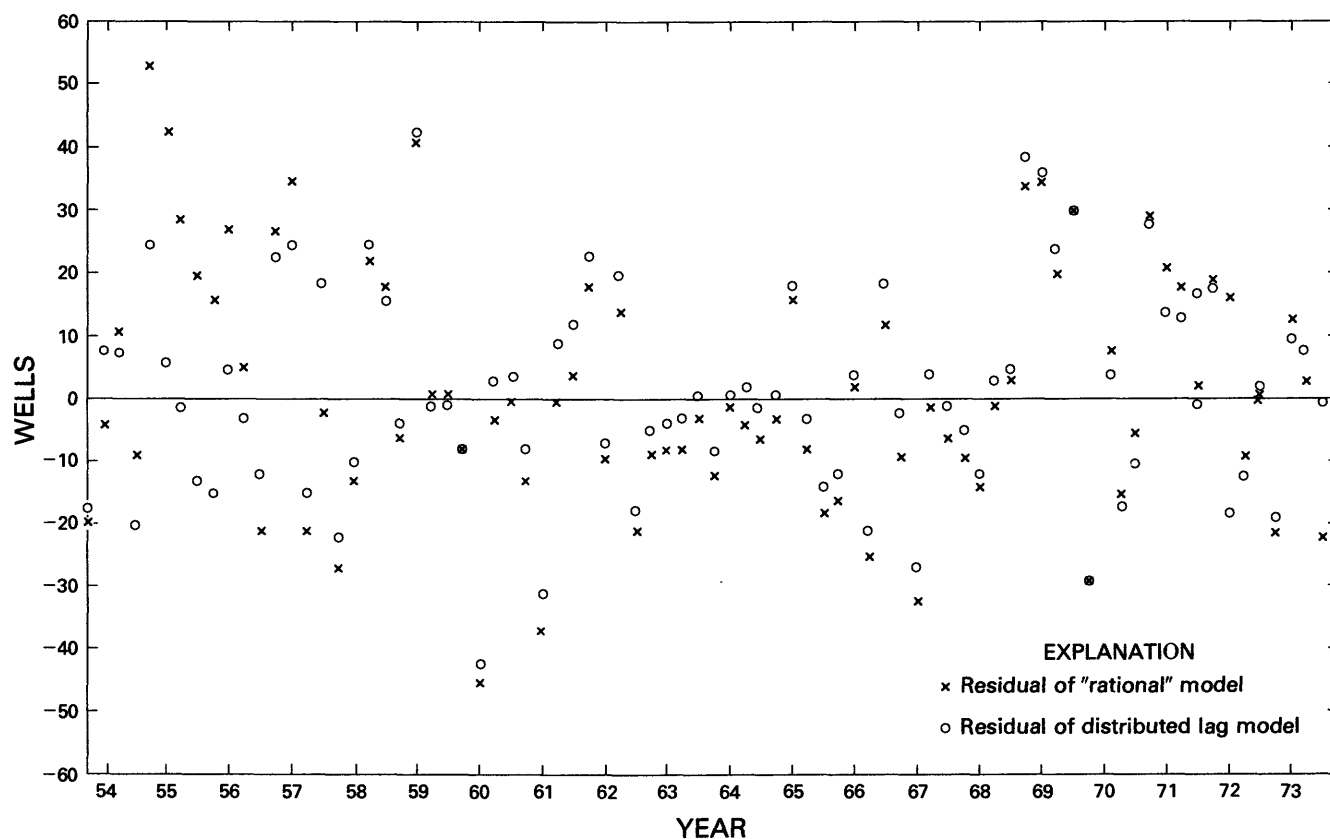


FIGURE 4.—Residuals for well-drilling models.

indicate systematic variation in either figure or for either model specification. The presence of a systematic component in the residuals would indicate that the model was misspecified. That is, if a significant variable was omitted from the specification, then a pattern in the residuals would correspond to systematic variations of the omitted variable. The most extreme residuals are accounted for by the "rational" model. Although the residuals vary widely for drilling from the third quarter of 1968 through 1970, the residuals for the drilling expenditure models indicate less erratic behavior.

Economic theory rarely specifies the functional forms that are used in applied econometric modeling. Consequently, in the process of model selection, several alternative functional forms are generally examined, and the final form is chosen on the basis of goodness of fit. This procedure is not entirely satisfactory, because specification errors resulting from incorrect or omitted variables are still possible. Specification errors may lead to biased estimates and predictions. A technique used to detect specification errors is to test for stability or structural changes in the model parameters during the original sample period (Jorgenson and others, 1970; Dhrymes and others, 1972). In particular, the full data sample is split into two subsamples or periods,

and the models are reestimated. The resulting parameters are tested to determine if statistically significant changes have taken place in their values.

The procedure used in detecting structural changes is the following: Suppose Q_2 is the sum of squared residuals associated with each subsample. Further, let Q_1 be the sum of squared residuals of the regression based on the pooled data. Under the hypothesis that the sets of regressors are equal for the two subperiods and the pooled sample, the test statistic is distributed according to the F distribution and is calculated (Chow, 1960) as

$$F = \frac{(Q_1 - Q_2)/k}{Q_2/(m+n-2k)} \quad (21)$$

with degrees of freedom $(k, m+n-2k)$. The variables m and n are the number of observations in each subsample, and k is the number of parameters that must be estimated in each model.

For the distributed lag models, the two subperiods were from the fourth quarter of 1952 to the first quarter of 1962 and from the second quarter of 1962 to the fourth quarter of 1973. The sample periods used for tests of the "rational" model were from the fourth quarter of 1953 to the third quarter of 1962 and from the

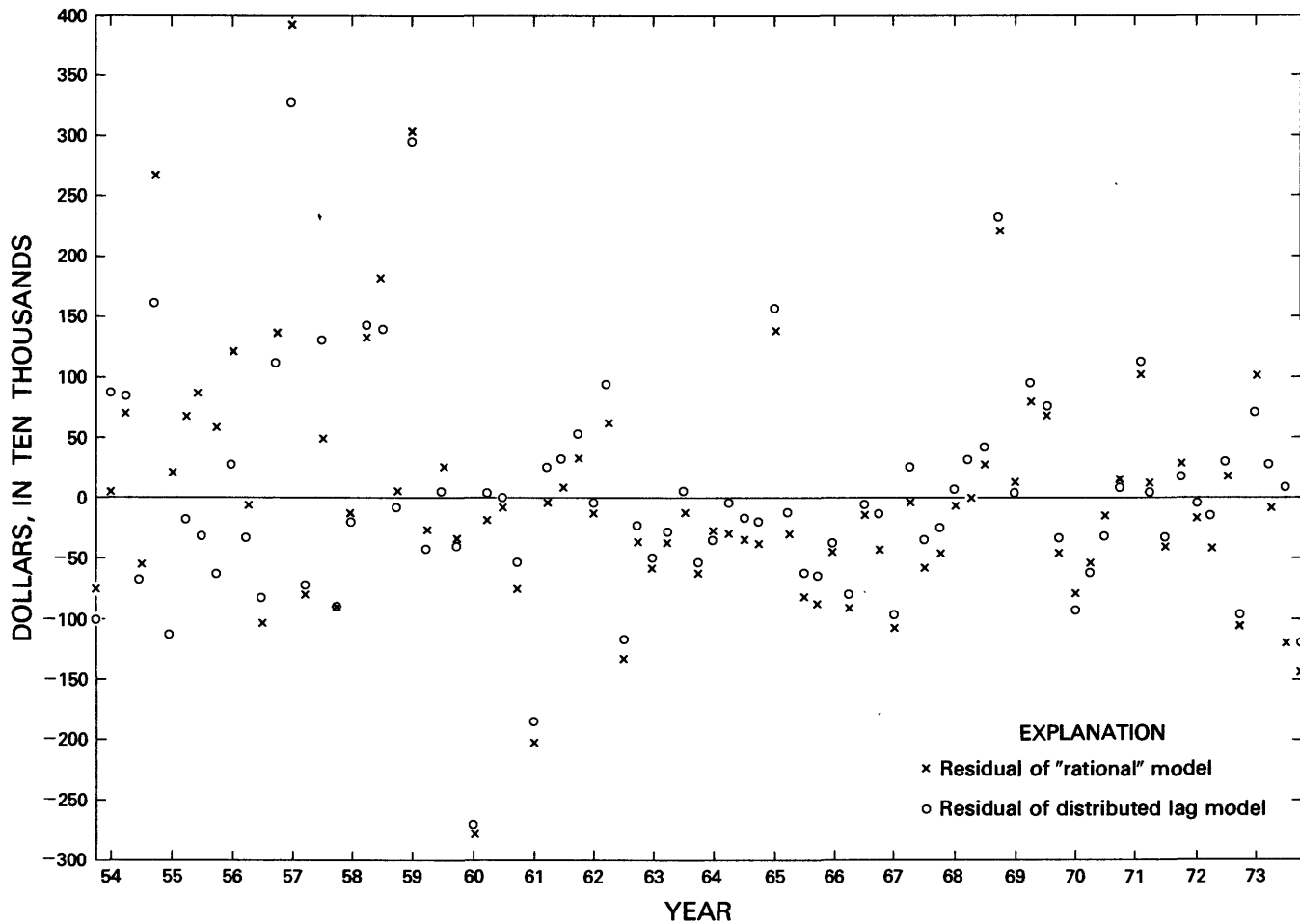


FIGURE 5.—Residuals for well-drilling expenditures models.

fourth quarter of 1962 to the fourth quarter of 1973. The level of type I error was set at 0.05; that is, one could expect to reject the null hypothesis about 1 time in 20 even if it were true. Rejection of the null hypothesis would indicate temporal parameter changes that might be the result of model specification error. For the distributed lag model the " F " statistics were 1.34 and 0.58 for the drilling and drilling expenditures equations, respectively. The model based on the predictions of the discovery process model had F statistics of 1.57 and 0.74 for the drilling and drilling expenditures equations, respectively. Thus there were no instances when the test statistic exceeded the corresponding critical level and the null hypothesis was not rejected, and the results of the tests for structural change provided no evidence that specification errors might be present in either set of models.

In conclusion, the test results comparing the distributed lag model and the model based upon predictions from the discovery process models indicated little difference in the potential predictive performance of the models. This finding is significant for three reasons.

First, it appears that the predictions of the discovery process model contain most, if not all, of the information included in the distributed lag functions of D_t and the experience-exhaustion variable k_t . Because of its relative simplicity, it may be more appropriate to use this model instead of the distributed lag model in applied situations. Second, if the discovery process model efficiently uses available information to generate predictions, then the specific distributed lag proxy used here does seem to capture operator expectations. Third, the reasonably high explanatory power of the "rational" model appears to indicate that the operators are using information contained in the discovery history of the basin efficiently and operating in a fashion consistent with these predictions.

SUPPLY OF RESERVES

The behavioral drilling models presented in the preceding sections appear adequate to explain the historical pattern of operator behavior in the basin. As such, they served as a basis for making inferences concerning the determinants of firm behavior. The estimated

behavioral models may also be applied in conjunction with the discovery process models to forecast the temporal pattern of new reserves, given certain assumptions about crude oil prices and drilling costs. The usefulness of forecasts of future reserves from undiscovered deposits is obvious. Furthermore, the generation of the forecasts by the analytical model allows the predictions to be conditioned on prices, well costs, and the current state of physical exhaustion. However, the accuracy of the forecasts is generally constrained by the nature of the models and the relevant behavioral and technical relations.

The estimated discovery process model mechanically provides predictions of the number of deposits within each size class as a function of the cumulative number of wildcat wells. To obtain a prediction of the reserves found for a given increment in wells, the field size (in barrels of oil) is multiplied by the incremental value in the predicted number of deposits. By linking the discovery process model to the behavioral drilling model, the predicted number of wildcat wells (per unit time) can be used to generate a forecast of the amount and value of reserves forthcoming. Because of the non-linear nature of the discovery process model, the linking of it and the behavioral drilling model must be relatively simple. Consequently, the integrated model is recursive in nature.

Although the discovery process model is independent of time, the behavioral drilling model is not. The drilling model was estimated with quarterly data, and consequently forecasts of the number of wildcat wells are also on a quarterly basis. Moreover, the model is specified so as to describe shortrun operator decisions. In general, the ability of econometric models to provide accurate forecasts deteriorates rapidly as the number of periods that the forecasts are made into the future increases. The extent of the deterioration can, to some degree, be determined by carrying out simulation experiments with the integrated equation system. In particular, these tests attempt to determine how rapidly the accuracy of the forecasts deteriorate, the robustness of the predictions of one endogenous variable when errors are in the other endogenous variable, and whether forecast errors are compensating or noncompensating as the results from individual time periods were aggregated.

A measure used to determine the quality of the integrated model's forecasting performance was the root-mean-square (RMS) prediction error (Thiel, 1966). It is defined as

$$\text{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - A_i)^2}, \quad (22)$$

where A_i is the actual value of the predicted variable and P_i is the prediction. Also used as measures of performance were the mean error (over several forecast periods) and the ratio of the RMS prediction error to the actual mean value of the variable predicted.

The integrated model is specified as a recursive equation system. Because of its simplicity, the "rational" well drilling specification was used rather than the distributed lag model. Variables that are endogenous or determined within the system are

y_t = number of wells drilled in period t and

R_t = reserves discovered in period t .

Variables that are taken as given or determined exogenously are

cw_t = cost per well in period t ,

p_t = price of crude oil, and

U_0 = the initial value of undiscovered deposits.

Among the predetermined variables that are taken as given are the initial values of the lagged endogenous variable y_{t-1} and $\hat{\pi}_{t-2}$. The discovery success ratio is not predicted, and consequently the profit variable was redefined to have the following form:

$$\hat{\pi}_{t-2} = \frac{R_{t-2} p_{t-2}}{y_{t-2}} - cw_{t-2}, \quad (23)$$

where R_t , p_t , cw_t , and y_t were defined as above. In the original estimation of the "rational" model, $F_i(u)$ was reestimated each period as new information became available or as discoveries were made. Reestimation of $F_i(u)$ each period for the integrated model would result only in the initialized value of the variable. Therefore, the final empirical estimates of the ultimate numbers of deposits for each size class were used. Because of the redefinition of the profit variable and the use of the final empirical estimates for $F_i(u)$, the "rational" well drilling equation was reestimated. The new equation is of the following form:

$$y_t = 8.386 + 0.06339 \hat{\pi}_{t-2} + 0.10358 \chi_t^* \quad (24)$$

(0.2) (1.7) (3.0)

$$\bar{R}^2 = 0.881 \text{ S.E.} = 21.1 \text{ Durbin-Watson statistic} = 1.53,$$

where the numbers in parentheses below the model coefficients are the " t " statistics, \bar{R}^2 is the adjusted coefficient of determination, and S. E. is the standard error of the regression equation. Using the Cochrane-Orcutt procedure to handle serial correlation, the first-order correlation coefficient was estimated to be 0.814. The predictive form of equation (24) (Kmenta, 1971) is

$$y_t = 0.814 y_{t-1} + (1 - 0.814) 8.386 + 0.06339 (\hat{\pi}_{t-2} - 0.814 \hat{\pi}_{t-3}) + 0.10358 (\chi_t^* - 0.814 \chi_{t-1}^*). \quad (25)$$

The predicted reserves found per time period, R_t , is calculated by taking the cumulative number of wells to period $t-1$, calculating the predicted (total) reserves, adding the number of wells drilled in period t to the cumulative well count, recalculating the predicted (total) reserves, and taking the difference between two total reserve predictions. The value of undiscovered deposits at the beginning of time $t+1$ was forecast by taking the difference between the predicted amount of discovered reserves at t and the sum of $F_i(u)$ across classes and multiplying the result by the crude oil price at the time. Predictions of y_t , R_{t-1} , and χ_{t+1}^* are employed in (23) and (25) to generate the predicted number of wells for period y_{t+1} , which in turn is used to predict R_{t+1} .

The historical period from the first quarter of 1960 to the fourth quarter of 1968 served as the base period for examining the integrated model's forecasting ability. This period was chosen because the estimated ultimate number of deposits in each size class had stabilized (Drew and others, 1980). The test period was cut off at the end of 1968 in order not to include the period when the Union Pacific Railroad acreage began to be drilled. We believed that the initiation of drilling on the new acreage represented an entirely exogenous influence, and the model could not be expected to reproduce it.

In order to determine how the accuracy of the forecasts was affected by the length of the forecast period, that is, how far ahead forecasts should be made, three sets of simulations were generated for the 9-year period. These include forecasts of 4, 8, and 12 periods (quarters) ahead. For example, during the 36 periods there were nine sets of 4-period forecasts. Using 8-period forecasts, four sets were generated, and for the 12-period forecasts, three sets were generated. Quarterly observations, particularly for new discoveries, have substantial stochastic components. Therefore, it is also of interest to determine if prediction errors during the forecast period offset each other when results for the time periods are aggregated.

Results of the experiments are presented in table 4. The table indicates the RMS prediction error, the mean (actual-predicted) prediction errors over the forecast time period, and the mean value of the actual variable that is being forecasted. The pattern of the prediction errors for wells indicates the degree by which lengthening the forecast period increases the RMS prediction errors and mean prediction error. The overall mean values of the RMS to the actual variable value are 0.141, 0.240, and 0.257, and the ratios of the value of mean period prediction error to the mean actual values are 0.033, 0.162, and 0.189 for the 4-, 8-, and 12-period forecasts, respectively. The obvious effect of increasing the forecast period is the accumula-

TABLE 4.—Forecast performance of integrated system

Forecast length ¹	Period	Wells			Reserves		
		RMS errors ²	Mean errors ³	Actual mean	RMS errors	Mean errors	Actual mean
A	1	7.551	3.589	109.250	2.103	0.786	3.763
	2	12.795	-9.213	92.250	3.183	-.025	3.161
	3	13.183	.868	110.000	3.036	-.573	4.111
	4	8.509	5.915	90.500	2.969	.277	2.555
	5	5.227	3.679	86.250	1.928	.330	2.096
	6	20.195	14.769	89.250	1.927	1.346	1.142
	7	12.703	9.698	68.750	3.381	-.519	2.374
	8	9.783	6.882	54.250	1.298	-.526	2.069
	9	19.146	-10.847	70.750	.585	-.458	1.837
Mean Values		12.121	2.816	85.694	2.268	.071	2.567
B	1	22.953	14.986	100.750	2.729	0.776	3.462
	2	19.434	12.399	100.250	2.987	.059	3.333
	3	13.424	5.803	87.750	1.915	.805	1.619
	4	28.126	23.671	61.500	2.549	-.275	2.221
Mean Values		20.984	14.215	87.563	2.545	.341	2.659
C	1	20.438	13.099	103.833	2.860	0.341	3.678
	2	15.643	11.746	88.667	2.345	.753	1.931
	3	29.878	23.761	64.583	2.102	-.141	2.093
Mean Values		21.986	16.202	85.694	2.436	.318	2.567

¹Forecast lengths A, B, and C are 4, 8, and 12 quarters, respectively.

²Root-mean-square value is given by $\sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - A_i)^2}$, where P_i is the predicted value and A_i is the actual value.

³Mean error is based on actual minus predicted value.

tion of prediction errors. It does not appear that the percentage errors increase linearly with the forecast period. To some degree, it appears that errors will be offsetting if the forecast period is taken to be sufficiently short, as evidenced by the relatively low mean error to actual value of 0.033 for the 4-period forecast simulation. As the forecast period is increased, the number of wells predicted appears to be systematically overestimated.

Results for the predicted amount of reserves found in new deposits are also presented in table 4. The discovery process model, used to predict the number of deposits and reserves forthcoming, produces relatively smooth estimates of forthcoming reserves. However, the actual number and sizes of new deposits, particularly for as short a period as a quarter, are highly erratic in nature. As a result, individual prediction errors as measured by the RMS prediction errors can be expected to be relatively large. However, if the model is operating reasonably well, in aggregating the predictions over time, errors will be compensating so that the mean prediction error should be relatively small. Results of the simulations appear to be consistent with this conjecture. The RMS prediction errors are very large, even for the set for which only 4-period forecasts were made. However, the mean prediction errors taken over the entire period are relatively small. The ratios of the mean prediction error values to the actual mean values are 0.028, 0.128, and 0.124 for the simulations based on the 4-, 8-, and 12-period forecasts, respectively. Individual errors in predicting forthcoming

reserves tend to offset each other as the errors are aggregated over time. The supply of reserves for increments in the drilling rate are too random to be predicted on a quarterly basis. Aggregation of quarterly forecasts to obtain annual values results in reasonably accurate predictions. The positive values of the mean forecast error for the 8- and 12-period simulation experiments does not necessarily mean that forecasts of the discovery process models are generally biased. The positive mean prediction error is probably the result of the overestimates for the number of predicted wells drilled as discussed earlier.

In summary, the prediction errors of the behavioral well-drilling model seem to be biased positively. However, the magnitude of the RMS and mean errors are quite comparable to the performance of well-drilling equations reported in other studies (MacAvoy and Pindyck, 1975) for similar forecast periods, that is, 4–8 periods ahead. The relatively high RMS errors for the additions to reserves appear to be more the result of the erratic or stochastic nature of the historical arrival of discoveries than systematic bias in the discovery process model. The relatively small mean prediction errors, particularly for the 4-period forecasts, indicate that if the quarterly forecasts of additions to reserves were aggregated, perhaps on an annual basis, and compared to the historical realizations, the RMS prediction errors would also be much smaller. The relatively small mean prediction error associated with the additions to reserves seems to suggest that the forecasts are reasonably robust even though there were errors in predictions of the number of wells to be drilled.

Although the linking of the discovery process model with the behavioral well-drilling model produced an analytical means of translating the forecasts of reserves per unit exploration effort, that is, wells drilled to reserves per unit time, several limitations of the analysis should be kept in mind. First, the economic model was specified to describe operator field decisions that are short term in nature. That is, the nature of decisions that are modeled are marginal adjustments in the rate of exploration rather than decisions to enter or exit a geologic basin. Second, because of the short-run nature of the models, it would be misleading to attempt to draw general conclusions about the effects of a general price change on drilling activity within the basin. That is, a general price change would induce some longrun adjustments to take place in the firm's internal allocation of resources across several regions. The type of price change that the behavioral well-drilling model might more appropriately capture corresponds to a change in the relative price of oil, for instance, the price of oil found in the Denver basin as

opposed to another basin. In order to predict the effects of a general change in the price of oil on drilling behavior for a particular basin, the behavioral well-drilling equation should be respecified to reflect the firm's long-run decisions and include a variable that would denote the firm's alternative exploration opportunities in other geologic basins or its alternative opportunities for obtaining additional reserves.

CONCLUSIONS

The purpose of this study was to examine operator exploration behavior at the field level. In the first part of the study an empirical model was specified and estimated. Distributed lag proxy variables were used to model operator expectations associated with the distribution of deposits remaining to be found in the basin. The estimated drilling models permitted inferences to be made about operator field behavior. First, if the form of the distributed lag function is interpreted to represent the operators' responses to new discoveries, then the nature of function is important. The appropriate form of the function was a finite polynomial lag function with an inverted "V" shape; that is, the effects of new discoveries increased for several periods and then diminished rapidly to 0 within 2½ years after discovery of the deposit. The independent variable used in the distributed lag model was the total value of deposits found in a given period with a minimum size of 500,000 barrels. Second, tests using alternative minimum size classes of deposits, that is, 2.5 million and 5 million barrels, in the distributed lag function produced no improvements in the fits. This result implies that, once an exploration play is underway, operator field expectations are sustained by the quantity of oil found per time period rather than by the discovery of specific size deposits. Comparison of these results and the regression coefficients showed them to be quite consistent with the results of a similar previous study based on the discovery history of the Powder River basin (Attanasi and Drew, 1977).

In the second part of the study, the discovery process model was applied to generate sequential predictions of the value of undiscovered deposits, which were used in the behavioral well-drilling model in place of the distributed lag operator-expectations proxy variable. The discovery process model efficiently uses information contained in the historical time series to estimate the future discoveries. Consequently, it was assumed that such predictions would closely correspond to "rational" expectations where "rational" is used in the sense that all available information in the historical data is used efficiently in the estimation process. Using the predictions from the discovery process model as the unobservable expectations variable associated with the

value of undiscovered deposits, the behavioral well-drilling equation was reestimated. The estimated "rational" model explained historical operator behavior well. This result seemed to indicate that operators, on the average, were efficiently using information contained in the discovery history of the basin and were behaving in a fashion consistent with these predictions. Comparison of the performance of the distributed lag model with that of the "rational" model indicated little differences in the predictive performance of the two models. The predictions of the discovery process model, used as "rational" expectations, appear to contain most if not all of the information included in the distributed lag functions and experience-exhaustion variable.

In the final section of the study, the behavioral well-drilling model and the discovery process model were integrated into a recursive equation system. The purpose of linking the models was to gain the ability to forecast additions to reserves in the time domain or on a unit time basis. The forecast performance of the integrated system was tested by examining the accuracy for various lengths of forecast periods. As might be expected, as the number of forecast periods increased, the accuracy of the forecasts deteriorated. However, the level of accuracy was still comparable to levels attained in other studies. Because of the stochastic nature of the discovery process, aggregation of quarterly predictions of additions to reserves to annual values tended to increase accuracy as measured by the mean error of prediction. The accuracy of the aggregated predictions appeared to be reasonably robust to the presence of prediction errors in the behavioral well-drilling equations. Finally, the integrated model provides for the first time a means of explicitly incorporating physical and economic exhaustion in the generated profit predictions and predictions of the value of remaining deposits into a behavioral model that describes the exploration effort (wells and expenditures) of operators.

REFERENCES CITED

- Almon, Shirley, 1965, The distributed lag between capital appropriations and expenditures: *Econometrica*, v. 33, no. 1, p. 176-196.
- American Petroleum Institute, 1953, 1955-56, 1959-73, Joint Association Survey of the U.S. Oil and Gas Producing Industry: Washington, D.C., American Petroleum Institute.
- Arps, J. J., and Roberts, T. G., 1958, Economics of drilling for Cretaceous oil on the east flank of the Denver-Julesburg basin: *American Association of Petroleum Geologists Bulletin*, v. 42, no. 11, p. 2549-2566.
- Attanasi, E. D., 1978, Firm size and petroleum exploration, Council of Economics of American Institute of Mining, Metallurgical, and Petroleum Engineers, Annual Meetings, Denver, Proceedings: p. 57-62.
- Attanasi, E. D., and Drew, L. J., 1977, Field expectations and the determinants of wildcat drilling: *Southern Economic Journal*, v. 44, no. 1, p. 53-67.
- Box, G. E. P., and Jenkins, G. M., 1970, Time series analysis: San Francisco, Holden-Day, Inc., 553 p.
- Chow, G. C., 1960, Tests of equality between sets of coefficients in two regressions: *Econometrica*, v. 28, no. 3, p. 591-605.
- Cox, J. C., and Wright, A. W., 1976, The determinants of investment in petroleum reserves: *American Economic Review*, v. 66, no. 1, p. 153-167.
- Dhrymes, P. J., 1971, Distributed lags: Problems of estimation and formulation: San Francisco, Holden-Day, Inc., 414 p.
- Dhrymes, P. J., Howrey, E. P., Hymans, S. H., Kmenta, Jan, Leamer, E. E., Quandt, R. E., Ramsey, J. B., Shapiro, H. T., and Zarnowitz, Victor, 1972, Criteria for evaluation of econometric models: *Annals of Economic and Social Measurement*, v. 1, no. 3, p. 272-328.
- Drew, L. J., Schuenemeyer, J. H., and Root, D. H., 1980, Petroleum-resource appraisal and discovery rate forecasting in partially explored regions; Part A, An application to the Denver basin: U.S. Geological Survey Professional Paper 1138-A, 11 p.
- Epplé, D. N., 1975, Petroleum discoveries and government policy: An econometric study of supply: Cambridge, Mass., Ballinger, 139 p.
- Erickson, E. W., and Spann, R. M., 1971, Supply response in a regulated industry: The case of natural gas: *Bell Journal of Economics and Management Science*, v. 2, no. 1, p. 94-121.
- Ezekiel, M. J., 1938, The cobweb theorem: *Quarterly Journal of Economics*, v. 53, no. 1, p. 255-280.
- Fama, E. F., 1972, Perfect competition and optimal production decisions under uncertainty: *Bell Journal of Economics and Management Science*, v. 3, no. 1, p. 151-174.
- Federal Energy Administration, 1974, Project independence blueprint: Final task force report: Oil: Possible levels of future production: Washington, D.C., U.S. Government Printing Office, 162 p.
- 1976, National energy outlook: Washington, D.C., U.S. Government Printing Office, 190 p.
- Fisher, F. M., 1964, Supply and costs in the U.S. petroleum industry: Baltimore, Johns Hopkins University Press, 177 p.
- Johnston, J. J., 1972, *Econometric methods* [2d ed.]: New York, McGraw-Hill, 437 p.
- Jorgenson, D. W., 1966, Rational distributed lag functions: *Econometrica*, v. 34, no. 1, p. 135-149.
- Jorgenson, D. W., Hunter, J. S., and Nadiri, M. I., 1970, The predictive performance of econometric models of quarterly investment behavior: *Econometrica*, v. 38, no. 2, p. 213-224.
- Kaufman, G. M., 1963, Statistical decision and related techniques in oil and gas exploration: Englewood Cliffs, Prentice-Hall, 307 p.
- Khazzoom, J. D., 1971, The FPC Staff's econometric model of natural gas supply in the United States: *Bell Journal of Economics and Management Science*, v. 2, no. 1, p. 51-93.
- Kmenta, Jan, 1971, *Elements of econometrics*: New York, MacMillan, 655 p.
- Leland, H. E., 1974, Production theory and the stock market: *Bell Journal of Economics and Management Science*, v. 5, no. 1, p. 125-144.
- MacAvoy, P. W., and Pindyck, R. S., 1973, Alternative regulatory policies for dealing with the natural gas shortage: *Bell Journal of Economics and Management Science*, v. 4, no. 1, p. 454-498.
- 1975, *The economics of the national gas shortage (1960-1980)*: Amsterdam, North-Holland, 259 p.
- National Petroleum Council, 1972, U.S. energy outlook: Washington, D.C., U.S. Government Printing Office, 260 p.
- Nerlove, M. L., 1958, Distributed lags and estimation of long-run supply and demand elasticities—theoretical considerations: *Journal of Farm Economics*, v. 40, no. 2, p. 301-311.

- Pesando, J. E., 1976, Rational expectations and distributed lag expectations proxies: *Journal of the American Statistical Association*, v. 71, no. 1, p. 36-42.
- Pierce, D. A., 1971, Fitting dynamic time series models—some considerations and examples: *Annual Symposium on the Interface Between Computer Science and Statistics*, 5th Stillwater, Okla., Proceedings, p. 1-2.
- Pindyck, R. S., 1978, Higher energy prices and the supply of natural gas: *Energy Systems and Policy*, v. 2, no. 2, p. 177-209.
- Root, D. H., and Schuenemeyer, J. H., 1980, Petroleum-resource appraisal and discovery rate forecasting in partially explored regions; Part B, Mathematical foundations: U.S. Geological Survey Professional Paper 1138-B, 9 p.
- Schmidt, P. J., and Waud, R. N., 1973, The Almon lag technique and monetary versus fiscal policy debate: *Journal of the American Statistical Association*, v. 68, no. 1, p. 11-19.
- Shiskin, Julius, Young, A. H., and Musgrove, J. C., 1967, The X-11 variant of the Census Method II seasonal adjustment program: Washington, D.C., U.S. Department of Commerce, 66 p.
- Thiel, Henrie, 1966, *Applied economic forecasting*: Amsterdam, North-Holland, 474 p.
- Turnovsky, S. J., 1969, A Bayesian approach to the theory of expectations: *Journal of Economic Theory*, v. 1, no. 3, p. 220-227.
- Uhler, R. S., 1976, Costs and supply in petroleum exploration—the case of alberta: *Canadian Journal of Economics*, v. 9, no. 1, p. 72-90.
- White, D. A., Gerrett, R. W., Marsh, G. R., Baker, R. A., and Gehman, H. M., 1975, Assessing regional oil and gas potential, in Haun, J. D., ed., *Methods of estimating the volume of undiscovered oil and gas resources*: Tulsa, American Association of Petroleum Geologists, p. 143-159.